SAND2011-3220C

A Neurophysiologically Inspired Hippocampus Based Associative-ART Artificial Neural Network Architecture

Craig M. Vineyard, Stephen J. Verzi, Michael L. Bernard, Shawn E. Taylor, Wendy L. Shaneyfelt, Irene Dubicka, Jonathan T. McClain and Thomas P. Caudell

Abstract—**Hippocampus within medial temporal lobe of the brain is essentially involved in episodic memory formation. Rather than simply being a mechanism of storing information, episodic memory associates information such as the spatial and temporal context of an event. Using hippocampus neurophysiology and functionality as an inspiration, we have developed an artificial neural network architecture called Associative-ART to associate k-tuples of inputs. In this paper we present an overview of hippocampus neurophysiology, explain the design of our neural network architecture, and present experimental results from an implementation of our architecture.**

I. Introduction

EPISODIC memory enables us to neurally encode personal experiences as represented from converged activations across cortical areas for diverse sensory modalities. In doing so, we are able to remember more than just a particular event. Rather, we are capable of remembering the detailed sequence of events comprising an experience as well as the temporal and spatial context of each event in the sequence [1]. One brain area, the hippocampus, is critically involved in remembering the spatial and temporal context of an event. Hippocampus location within human brain may be seen in Fig. 1. The medial temporal lobe (MTL), where hippocampus is located, is the recipient of inputs from widespread areas of the cortex and supports the ability to bind together cortical representations [1].

Manuscript received February 10, 2011. This research was possible in part by LDRD program support from Sandia National Laboratories. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

C. M. Vineyard is with Sandia National Laboratories, Albuquerque, NM 87185 USA (phone: 505-284-4118; fax: 505-844-8604; e-mail: cmviney@ sandia.gov).

S. J. Verzi is with Sandia National Laboratories, Albuquerque, NM 87185 USA (e-mail: sjverzi@sandia.gov).

M. L. Bernard is with Sandia National Laboratories, Albuquerque, NM 87185 USA (e-mail: mlberna@sandia.gov).

S. E. Taylor is with Sandia National Laboratories, Albuquerque, NM 87185 USA (e-mail: setaylo@sandia.gov).

W. L. Shaneyfelt is with Sandia National Laboratories, Albuquerque, NM 87185 USA (e-mail: wlshane@sandia.gov).

I. Dubicka is with Sandia National Laboratories, Albuquerque, NM 87185 USA (e-mail: idubick@sandia.gov).

J.T. McClain is with Sandia National Laboratories, Albuquerque, NM 87185 USA (e-mail: jtmccl@sandia.gov).

T. P. Caudell is with the University of New Mexico, Albuquerque, NM 87131 USA (e-mail: tpc@ece.unm.edu).

A key component of episodic memory is association formation [2]. This capability allows you to relate knowledge pertaining to elements of an event such as who, what, when, and where. In this paper we present an artificial neural network architecture for association formation inspired by hippocampal functionality. First we give a brief overview of hippocampal neurophysiology, and then we provide an explanation of how key hippocampal functionality is incorporated into the presented architecture. Next we provide experimental results from an implementation of the architecture, and then close with conclusions and future work.

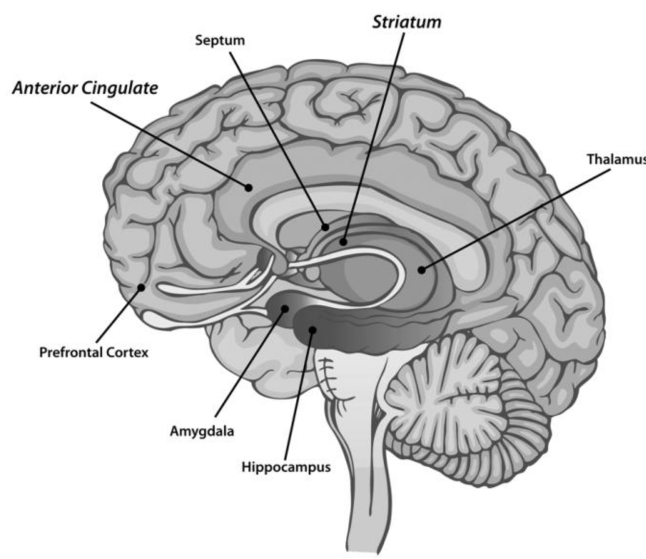

Fig. 1. Hippocampus location within human brain.

II. Hippocampus Neurophysiology

Cortical inputs to MTL arrive from various sensory modalities, with different emphases depending upon the mammalian species. For instance, rats receive a significant olfactory influence whereas bats receive a strong auditory influence [3]. Nevertheless, across species, most of the neocortical inputs to the perirhinal cortex come from cortical areas which process unimodal sensory information about qualities of objects ("what" information), and most of the neocortical inputs to the parahippocampal cortex come from cortical areas which process polymodal spatial ("where") information [3][1]. There are some connections between the two streams, however overall processing of the streams remains largely segregated until they converge within hippocampus [5].

Extensive neuroscience research typically identifies hippocampus to be composed of a loop receiving inputs from entorhinal cortex (EC), which receives inputs from perirhinal and parahippocampal cortices, and beginning with dentate gyrus (DG), proceeding to CA3, followed by CA1 and propagating back to cortex. These sub regions will be addressed individually as follows.

The DG receives the conjoined multimodal sensory signals from EC. Anatomically, DG consists of a large number of neurons with relatively sparse neural activation at a given instant. Effectively, this behavior suggests that the DG creates non-overlapping sparse codes for unique events [6]. The sparse DG outputs serve as the input for CA3.

The CA3 region of hippocampus consists of extensive recurrent connections. Additionally, the presence of numerous inhibitory and excitatory interneurons enables CA3 to perform auto-association processes. Anatomically, the output of CA3 proceeds to CA1 and subiculum as the major output regions of hippocampus [7].

While the exact functionality of subiculum is largely unknown, CA1 functionality is typically identified as learning relational information for temporal sequences and connecting episodic encodings from CA3 with the original EC sensory activations.

We have used some of these functional properties of hippocampus as the basis for an artificial neural network architecture for association formation which we will describe next.

III. Computational Architecture

In general, an association is a relationship between entities of a particular type. For example, an individual is associated with their name or two individuals may be associated by a common workplace. All entities are trivially related to themselves, but more interesting associations are between pairs and k- tuples of entities. A pair is the simplest non trivial association, but more complexly, k individual entities may be associated with each other as a k-tuple. And so the question arises as to how relationships are formed.

Numerous domain specific rules or heuristics may be derived based upon criteria such as distance metrics or shared features. But instead, our architecture, which is inspired by hippocampus, answers this question by the premise of associating a focus with its context, analogous to the dorsal and ventral partitioning in EC sensory input signals. In other words, our approach associates what and where information based upon their shared frame of reference. For example, a man may be associated with the home he is seen living at.

However, beyond simply deciding what entities should be associated with one another there is also the issue of representation. Prior to entering hippocampus, sensory signals pass through numerous layers of cortex. Throughout these layers a representation for entities are built up. Eventually, within hippocampus, the DG is believed to create unique sparse encodings for unique perceptions. Likewise, our architecture relies upon having a unique representation of the inputs it receives such that it can identify whether the current input is an item it has seen before and update any existing associations appropriately, or whether the input is novel necessitating a new encoding.

Our architecture, shown in Fig. 2, addresses this capability by using fuzzy-Adaptive Resonance Theory (ART) artificial neural network modules. Developed by Carpenter and Grossberg, the ART family of neural networks are online, unsupervised neural networks which are excellent at category formation [8]. The fuzzy-ART variant which we have employed in our architecture operates upon real valued inputs. Given a vector of real valued numbers corresponding to a particular input, fuzzy-ART performs pattern categorization and through winner-take-all competition yields a unique output value to represent a group of similar inputs. A vigilance parameter allows us to control how similar inputs are to be grouped within the same category. A vigilance value of one specifies that the inputs must be identical. Lowering the vigilance parameter towards zero allows for generalization such that similar, but not exactly identical, inputs may be grouped together. If no existing category is sufficiently close to represent a novel input, then ART is capable of expanding and creating a new category. We have utilized these capabilities by employing a fuzzy-ART module to categorize the inputs presented to our architecture. In the neurophysiology, DG creates nearly unique encoding for novel inputs. Likewise, the fuzzy-ART module we are using in our architecture creates representative categories for inputs. Repeated presentation of previously seen (identical) inputs activates the same categorical representation whereas newly seen inputs can be represented by their own encoding. These unique categorical activations may then be further processed and associated together.

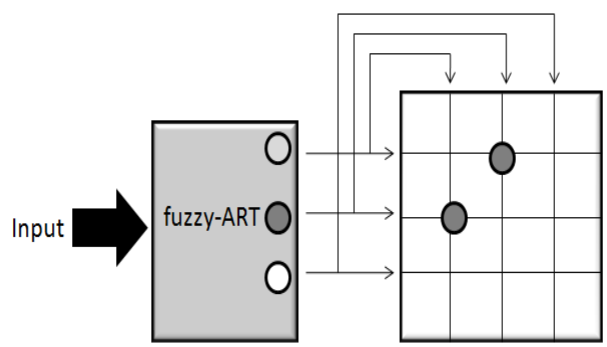

Fig. 2. Associative-ART Architecture.

The DG encodings of hippocampus propagate to the CA3 region which is believed to be heavily composed of recurrent connections and associations. In our architecture, by connecting an association field to the template activations of the fuzzy-ART module we are able to encode associations among k-tuples of inputs. Existent neural network architecture ARTMAP links two ART modules using a mapfield such that the mapfield may record simultaneous activations across the two ART modules. The rectangular mapfield of ARTMAP connects one ART module to each axis of the map grid and the intersecting grid lines encode a connection between the two ART modules [9]. The ARTMAP architecture allows many-to-one associations to be formed from ART_a to ART_b where the a-side ART module receives input from a data vector and the b-side ART module receives input from a label vector in (supervised learning) classification tasks (see Figure 1 in [9]).

Our Associative-ART architecture consists of only a single fuzzy-ART module and utilizes the association field to encode associations between k-tuples of entities presented to it rather than between two ART modules. Instead of connecting a separate ART module to each axis of the association field, the outputs of our single ART module are mirrored connecting it to both sides of the association field and subsequently allowing associations to be formed across the single ART module. All association field values are initialized to zero. Upon receiving a k-tuple input, associations are formed by handling all pairs. For each of the pairs, the grid intersection of the two entities in the association field A is set to a value of one as shown in the following equation:

$$A_{ij} = 1, \forall (i, j) \in P, \tag{1}$$

where P is the set of all pairs of elements in a k-tuple input. Each element in the k-tuple input will correspond to a particular fuzzy ART category activated during the previous k time steps. Consequently, the association field of our Associative-ART architecture creates a symmetric binary association matrix.

The overall Associative-ART architecture is depicted in Fig. 2. Using a single fuzzy-ART module necessitates that rather than presenting associated inputs simultaneously, they are presented sequentially to the fuzzy-ART module within the architecture. Rather than encoding the instantaneous activation of an individual input, the association field associates the previous k fuzzy-ART outputs. In other words, a single association field update encompasses k fuzzy-ART categories. There is no sequencing in the association field; instead there are multiple simultaneous activations as may be seen in Fig. 2. In the Fig. 2 example, the association formed links the dark gray and light gray categories as a paired association in the association field. Associations are symmetric and may be many-to-many.

IV. Image Association Experiment

As a demonstration of the associative capability of our Associative-ART architecture we have constructed a simple image association experiment with 13 unique inputs and 14 associations amongst the inputs. The parametric configuration we used for the fuzzy-ART module is β set to 1 (fast learning), a choice parameter α of 0.01, and a vigilance of 0.99. As the base case, we have set k equal to two so that the associations are pairs. While ART is capable of processing any vectorized inputs, for this experiment we have presented our architecture with images of uniquely numbered circles as shown in Fig. 3. Each row in the figure portrays an associative pairing and the column depicts the individual input which was presented to the architecture.

V. Image Association Results

In the association field of our Associative-ART architecture, the ordering of the pairs is arbitrary in regards to the overall result. However, computationally by using a fuzzy-ART neural network, the ordering influences the representative template encoding of the input. For example, Input B in the first row is the same image as Input A in the second row. Due to the fact that both Input A and Input B are processed by the same Fuzzy-ART module, the repeated presentation of an input is represented by the same output activation as opposed to a unique encoding whether the input was presented as Input A or Input B.

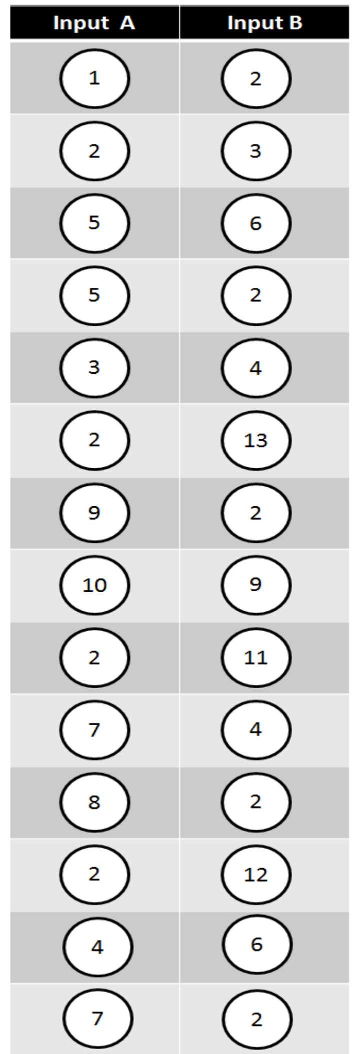

Fig. 3. Input image pairs for experiment testing association capabilities of the architecture.

For this simple example, we were able to manually construct the association field generated by the pairing of the inputs for comparison purposes and verify the association field generated by our Associative-ART architecture was equivalent. The association field generated by our architecture is shown in Fig. 4, which is identical to the manually constructed association field. In this representation, each row of the association matrix represents an ART category corresponding to a unique input. Likewise, the columns are a mirror of the rows. And so, a value of one for a particular row-column intersection denotes an association between the respective inputs represented by those ART categories. Zeroes in the association matrix depict no known association between two entities, and have been omitted from Fig. 4 for clarity. For undirected associations, as is the case in this example, the resulting association matrix is symmetric about the main diagonal.

	1	2	3	5	6	4	13	9	10	11	7	8	12
1		1											
2	1		1	1			1	1		1	1	1	1
3		1				1							
5		1			1								
6				1		1							
4			1		1						1		
13		1											
9		1							1				
10								1					
11		1											
7		1				1							
8		1											
12		1											

Fig. 4. Graph of resulting associations.

Additionally, as a more intuitive but equivalent depiction, we have generated an association graph from the association field which is shown in Fig. 5. As illustrated in this figure, while simple pairs were presented to the architecture, the net result is a more complex associative graph or network in which larger transitive and group associations may be inferred. For example, while input circles 10 and 4 were never associated with one another, transitive association paths exist by which the two inputs may be connected.

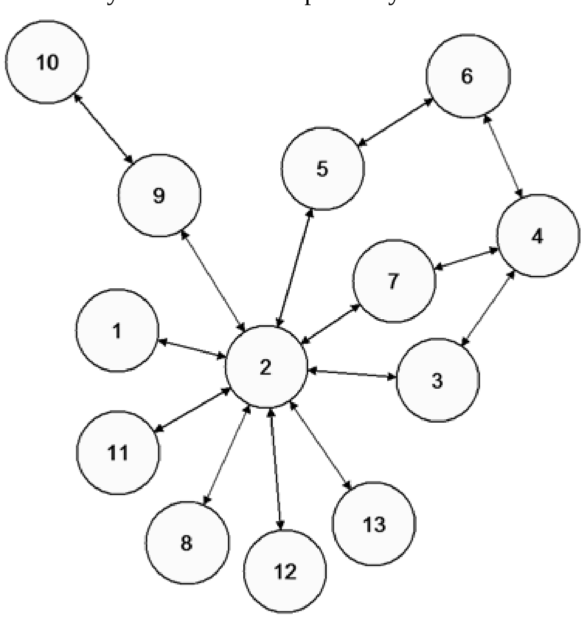

Fig. 5. Graph of resulting associations.

VI. Text Association Experiment

As a second exemplar illustrating the associative capabilities of our architecture, we have created a text based association example. Just as an individual's physical self is associated with their name, an individual's first and last name are associated with one another. And so, for this example we have used the first and last names of United States Presidents as our dataset [10]. First and last names were presented to the architecture individually and consecutively as text strings. John Quincy Adams and the two President Bush's were presented as triples rather than pairs to differentiate these individuals. The full input data is portrayed in Fig. 6. The ordering shown in the image is the same as the order presented to our architecture, and although the ordering does not affect the final associations formed we have presented the paired names in order of presidency.

In order to process text strings using a Fuzzy-ART module the text string must be mapped to a numeric vector. In this case, we have done so by forming a vector consisting of the American Standard Code for Information Interchange (ASCII) decimal value for each individual letters in the names [11]. Additionally, because Fuzzy-ART requires a fixed length input vector, we have padded the shorter names with zero values at the end of the vector to attain a constant length for all text strings. Alternative numeric text encodings are possible but would not alter the resulting association formations. The parametric configuration we used for the fuzzy-ART module is β set to 1 (fast learning), a choice parameter α of 0.01, and a vigilance of 1.

VII. Text Association Results

This example illustrates our architecture's ability to operate upon various input types, not just graphics as demonstrated in the first example. The particular characteristics of an association graph or network are dependent upon the data presented to the architecture. This second example is a larger input data set which exhibits some characteristics not present in the first example. Due to the increased complexity of this example, the full association matrix generated by the architecture is too large to meaningfully display within this paper. However, Fig. 7 depicts a few of the interesting associations extracted from the overall resultant association matrix.

In this example, the overall association graph is not connected, but rather disjoint groupings form, some of which are shown in Fig. 7. For example, as may be seen on the right side of Fig. 7, James has been a popular first name among several presidents and thus six different presidents are all associated with this first name. Other associations are unique in the sense that they do not share a first or last name with any other president and consequently only the two names are associated with each other and nothing else. Two such examples shown in Fig. 7 are Abraham Lincoln and Ronal Reagan near the top right corner of the figure.

John Adams was the second President and John Quincy Adams was the sixth President. As previously stated, the name John Quincy Adams was presented as a triple to differentiate these two men.

First Name:	Last Name:
George	Washington
John	Adams
Thomas	Jefferson
James	Madison
James	Monroe
John Quincy	Adams
Andrew	Jackson
Martin	Van Buren
William	Harrison
John	Tyler
James	Polk
Zachary	Taylor
Millard	Fillmore
Franklin	Pierce
James	Buchanan
Abraham	Lincoln
Andrew	Johnson
Ulysses	Grant
Rutherford	Hayes
James	Garfield
Chester	Arthur
Grover	Cleveland
Benjamin	Harrison
Grover	Cleveland
William	McKinley
Theodore	Roosevelt
William	Taft
Woodrow	Wilson
Warren	Harding
Calvin	Coolidge
Herbert	Hoover
Franklin	Roosevelt
Harry	Truman
Dwight	Eisenhower
John	Kennedy
Lyndon	Johnson
Richard	Nixon
Gerald	Ford
James	Carter
Ronald	Reagan
George H.W.	Bush
William	Clinton
George W.	Bush
Barack	Obama

Fig. 6. Input text pairs of U.S. President Names.

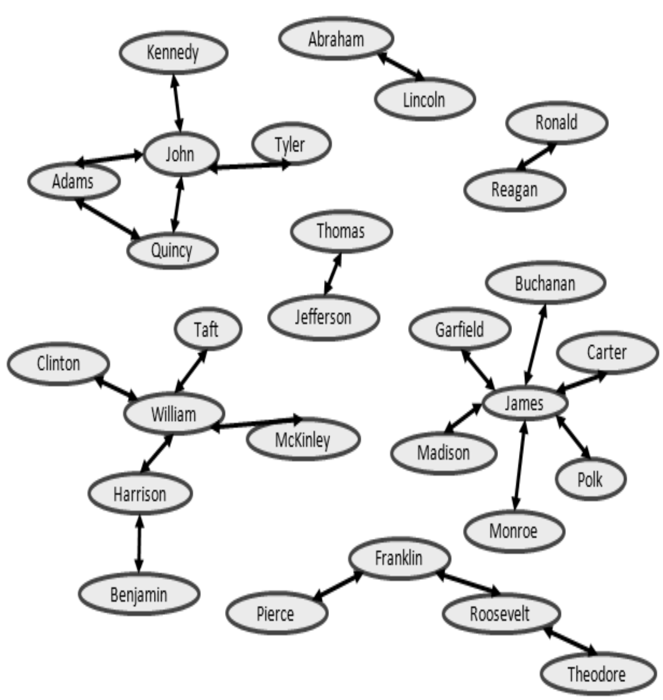

Fig. 7. Partial graph of text associations of U.S. President Names.

The graph of the resulting association group is illustrated in the upper left portion of Fig. 7.

VIII. Conclusions and Future Work

In this paper we have presented an artificial neural network computational architecture with functionality inspired by the neural functionality of hippocampus. Specifically, this architecture was designed based upon the DG and CA3 regions of hippocampus to learn associations amongst k-tuples of entities. Our approach is general, as opposed to a domain specific solution, in the sense that it can handle any sort of input as long as the input may be represented as a numeric vector.

In this paper we have demonstrated the architecture on a couple of simple problems which begin to show the architecture's potential for representing non-explicit association networks. Constructing association networks such as these allows further analysis such as transitivity, centrality, clustering, connectivity, and other network metrics. Additionally, in regards to data mining, our approach provides a means of representation and structured presentation.

Future development of this architecture may include additional processing within the association field. Rather than simply recording a binary association value, additional metrics such as a frequency count or a recency value may provide interesting enhancements. Incorporating a frequency count is one possibility to identify strength of association such that pairings repeatedly presented together are more strongly associated than items only presented once. In our preliminary architecture, presentation order is irrelevant, but if instead order matters a temporal marker could be utilized to assess how recently an association was formed. From this approach, various further processing could be incorporated such as the decay of associations over time. Depending upon the particular application, architecture modifications such as these provide great potential for enhanced further processing as well as addressing episodic or sequential data.

Acknowledgment

This research was possible in part by LDRD program support from Sandia National Laboratories. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

[1] H. Eichenbaum, "The hippocampus and declarative memory: cognitive mechanisms and neural codes," *Behavioural Brain Research*, 2001, pp. 199-207

[2] Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., Nash, C. (1999). Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus, 9:* 83-98.

[3] R.D. Burwell, M.P. Witter, & D.G. Amaral. Perirhinal and postrhinal cortices of the rat: a review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus*, (1995), 390-408.

[4] W.A. Suzuki, & D.G. Amaral,Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *Journal of Comparative Neurology 350*, (1994), 497-533.

[5] Suzuki, W., & Eichenbaum, H. (2000). The neurophysiology of memory. *Annals of the NY Academy of Sciences*, *911*, 175-91.

[6] Leutgeb, et al, Independent codes for spatial and episodic memory in hippocampal neuronal ensembles, *Science* 309, (2005).

[7] Eichenbaum, H. & Cohen, N.J. (2001). From Conditioning to Conscious Recollection: Memory Systems of the Brain. Oxford: Oxford University Press.

[8] G.A. Carpenter, S. Grossberg, and D.B. Rosen. FuzzyART: fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, (1991), 4:759-771.

[9] G.A. Carpenter, S Grossberg, N Markuzon, J.H. Reynold, & D.B. Rosen. Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, (1992), 3: 698-713.

[10] "The Presidents" http://www.whitehouse.gov/about/presidents

[11] "ASCII Table and Description" http://www.asciitable.com/