SAND2010-3512C

A Cortical-Hippocampal Neural Architecture for Episodic Memory with Information Theoretic Model Analysis

Craig Vineyard, Shawn Taylor, Michael Bernard, Stephen Verzi, Tu-Thach Quach
Sandia National Laboratories, PO Box 5800 Mail Stop 1188,
Albuquerque, NM 87185 1188, United States

and

Thomas Caudell
Electrical and Computer Engineering Department, University of New Mexico, MSC01 1100, 1 University of New Mexico, Albuquerque, NM 87131-0001, United States

and

Patrick Watson
Beckman Institute, University of Illinois at Urbana-Champaign, 405 North Mathews Avenue, Urbana, Illinois 61801, United States

ABSTRACT

Extensive neuroscience research on the hippocampus has identified its crucial role in memory formation and recall. Specifically, associative binding of the components comprising an episodic memory has been identified as one of the functions performed by the hippocampus. Based upon neuroanatomical function we have devised a computational cortical-hippocampal architecture using variants of adaptive resonance theory (ART) artificial neural networks. This computational model is capable of processing multi-modal sensory inputs and capturing qualitative memory phenomena such as auto-association and recall. Model performance is assessed both qualitatively and quantitatively. From a quantitative standpoint, we have applied the mathematics of information theory to quantify the similarity between recalled images yielded by the model and the unaltered original inputs. Thus in this paper we present a neurologically plausible computational architecture as well as a quantitative assessment of model performance.

Keywords: Artificial neural network, hippocampus, information theory, computational neural architecture

1. INTRODUCTION

Although tragic, legendary neuroscience patient H.M.'s memory impairment was instrumental in instigating awareness of the role played by medial temporal lobe and specifically the hippocampus [6]. Rather than serving as a central repository of memories, it was recognized that the hippocampus is key to forming episodic memories. Furthermore, associative binding of the components comprising an episodic memory is one of the functions performed by the hippocampus.

Modeled after cortical-hippocampal structure and function, this paper presents a neural architecture for episodic memory formation and recall developed as a collaborative effort by Sandia National Laboratories, the University of Illinois at Urbana Champaign, Boston University, and the University of New Mexico.

2. MODEL ARCHITECTURE

We have devised a computational cortical-hippocampal architecture using variants of adaptive resonance theory (ART) [1] artificial neural networks as the fundamental component. Our architecture is guided by accepted hippocampus sub-region functionality, neural density, and connectivity. To this effect, as the entry point to hippocampus, our representation of the entorhinal cortex (EC) facilitates the convergence of multiple sensory streams by uniting dorsal and ventral visual streams. This combination is then received by our representation of dentate gyrus (DG) region which performs a kind of pattern reduction and separation. Maintaining an approximate anatomically correct ratio of neural inputs to outputs, we have represented dentate gyrus as a series of winner-take-all fuzzy-ART modules. This effectively yields sparse encodings for differing input signals. Resultant unique representations from the DG module are auto-associated within CA3 (in cornu ammonis) such that related memories are bound together. Computationally, this is implemented by incorporating self-organizing map neighborhood update properties within the learning rules of an ART module. And finally the major output regions

Deleted:

of the hippocampal loop, a conjoined representation of CA1 (also in cornu ammonis) and subiculum creates a temporal sequence linking of CA3 encodings back to the original entorhinal cortex representation. This encoding is represented as a semi-supervised laterally primed adaptive resonance theory module.

Our cortical-hippocampal architecture is depicted in Figure 1, portraying the incorporated modules as well as their connectivity and data flow. As illustrated in Figure 1, beginning at the bottom, two separate processing streams focus upon different information sources but converge as they move upwards to the hippocampal loop representation (top part of Figure 1). The convergence of information streams proceeds to propagate through the hippocampus and back out to cortex for long term storage.

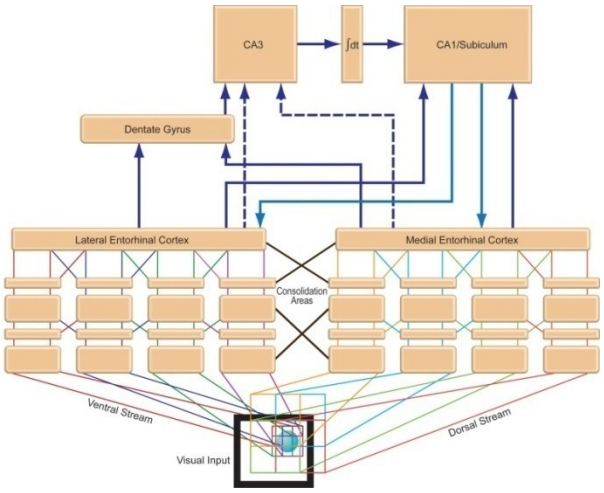

Figure 1: Architectural Overview

3. QUALITATIVE ANALYSIS

This computational model is capable of processing multi-modal sensory inputs and capturing qualitative memory phenomena such as auto-association and recall. Neuroscience research has identified that the process of neural activation within the hippocampal loop propagates from EC to DG, to CA3, and finally out through CA1/subiculum back to EC. From a qualitative standpoint, as a baseline comparison we compare neural activations within our model to ensure they exhibit the same flow. Additionally, we have also compared performance with human subjects in the ability to automatically associate novel relationships between visual stimuli that have a shared context but are never explicitly shown together. This capability may be qualitatively perceived by observing the similarity in activations within our model. Figures 2 and 3 depict the graphical user interface (GUI) of our model in which the input images are captured in the lower left hand box, and the remaining regions portray activation within the various neural regions. Of particular interest is the CA3 activation in the upper left. As shown in the figures, when distinctly different faces are paired with a common house, this notion of cohabitation is captured through CA3 association encoding. For a more in depth description see [8].

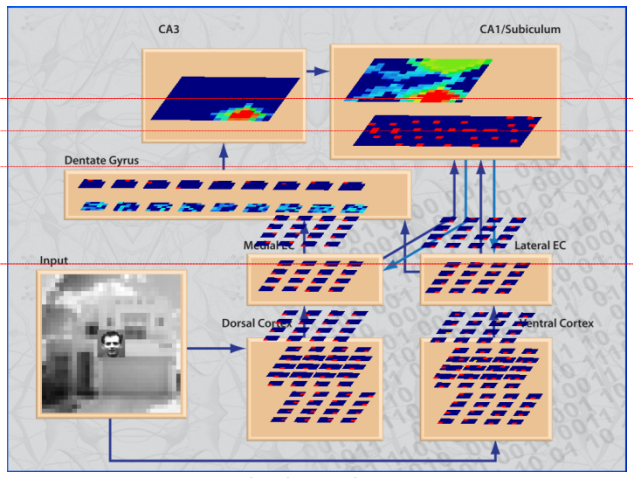

Figure 2: Novel Relationship Association 1

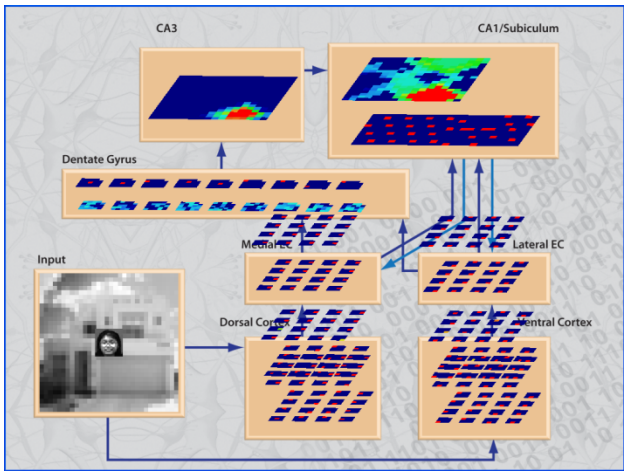

Figure 3: Novel Relationship Association 2

4. COMPUTATIONAL EXPERIMENT PARADIGM

In addition to qualitatively assessing phenomena such as activity of particular neural regions and data flow paths, we have compared our model's performance with that of human subjects in experiments conducted by other researchers. In particular, we have compared the model to a study performed by Preston et al. in which human subjects are trained on black and white photographs of face-house pairs [4]. Following the training phase, during testing subjects performed a forced-choice judgment task in which they were shown only either a face or house and were required to answer which corresponding house or face completed the pair. Additionally, they were also tested on their ability to

identify related face-face pairs which were independently presented in conjunction with a common house, but were never seen together.

We have replicated this procedure by presenting to our model low resolution face-house pairs, such as that shown in Figure 4. Analogous to the forced-choice judgment task, we presented the model a partial input cue by presenting a blank as one of the inputs, and subsequently the model performs a full recall reconstructing the image associated with the non-blank input. This experimental paradigm may be seen as follows in Figure 5.

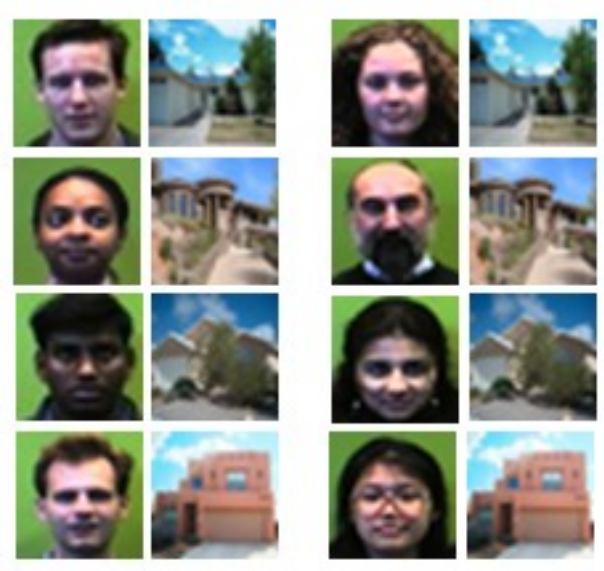

Figure 4: Input Face-House-Pair Images

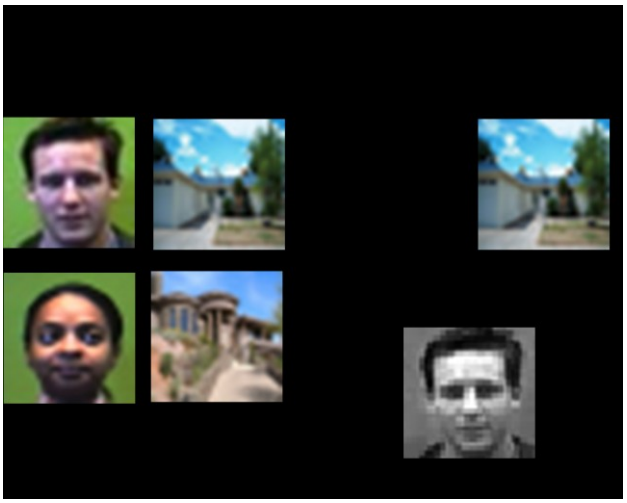

Figure 5: Forced-choice Judgment Computational Paradigm

5. QUANTITATIVE ANALYSIS

In the human subject realm, in addition to neural activation, a verbal response is given to gauge learned associations. We have not modeled all of the appropriate neural and physiological mechanisms to directly compare our model's output with human subject verbal responses. However, the classic Turing test seeks to equate the playing field when comparing humans and computers by abstracting the traits which are inherently human or machine [7]. In a similar manner, just as the human subjects are not required to reproduce a hand drawn image of their memory but rather select the choice closest to their memory, we do not expect our model to produce a perfect recall but rather interpret its recollection to be the image which is closest to its recalled encoding. Assessing which image is closest provides a means of quantitatively analyzing our model.

Although a seemingly simple concept, distance may be assessed in many different ways. The Euclidean distance between two points in a plane measures how close they are in space. Hamming distance defines the distance between two binary strings of equal length as the number of positions in which the two strings differ. It quantifies the number substitutions required to make two strings match. One of the most significant applications of Kolmogorov complexity is the normalized information distance [3]:

$$NID(x,y) = \frac{\max\{K(x|y),K(y|x)\}}{\max\{K(x),K(y)\}} \quad (1)$$

However, a major drawback of Kolmogorov complexity is the fact that it is non-computable. That is, given x, we cannot compute $K(x)$ exactly. Rather, we can only approximate $K(x)$. As a consequence, we cannot compute the NID directly. Instead, this metric may be approximated using compression. Given a compressor C, the resulting approximation of the NID is the normalized compression distance [2],[3],[5]:

$$NCD(x,y) = \frac{C(x,y)-\min\{C(x),C(y)\}}{\max\{C(x),C(y)\}} \quad (2)$$

where $C(x)$ is the compressed size of x and $C(xy)$ is the compressed size of the concatenation of x and y.

Using the Lempel-Ziv-Markov chain algorithm (LZMA) for compression, we were able to compute the NCD of two images. And so, without the model being able to provide explicit output this yields a quantifiable way of assessing model performance.

6. RESULTS

We have performed two experiments with our model using the eight face house pairs portrayed in **Error! Reference source not found.** The pairs were presented sequentially one pair at a time presenting the data in the left column followed by that of the right. Faces associated with a shared house were uniformly spaced, in the presentation sequence, with three different

face-house pairs interjected in between. Additionally, opposite genders were paired with a given house such that the association formation was based upon shared context and not similarity in features.

The results of the first experiment using this data set are recorded in Table 1. Contained within this chart are the pair wise NCD values for a particular face (across the row) against the four possible houses (columns). The smallest NCD value then corresponds to the correct answer (or house). As shown, the model was correct on six of eight presentations. House1, represented by the first column, is never selected as the recalled house. This missing representation led to the two erroneous answers, Face1 and Face5, which should have been associated with House1. Rather, in each case they were coupled with another one of the faces with similar features. For example, Face5 has some perceivably similar facial features as Face7, whom is appropriately paired with House3 (which Face5 answered).

	House1	House2	House3	House4
Face1	0.849	0.8571	0.8601	0.2308
Face2	0.85	0.4104	0.8599	0.8268
Face3	0.8432	0.861	0.5191	0.8484
Face4	0.849	0.8571	0.8601	0.2308
Face5	0.8432	0.861	0.5191	0.8484
Face6	0.85	0.4104	0.8599	0.8268
Face7	0.8432	0.861	0.5191	0.8484
Face8	0.849	0.8571	0.8601	0.2308

Correct Answer | Incorrect Answer

Table 1: Preliminary Experimental Results

To address the issue of this misrepresentation, we adjusted the representational fidelity of the model. ART neural networks include a vigilance parameter which dictates the precision of category representation [1]. By using ART neural networks as the fundamental component of the architecture we can alter the precision of the model by adjusting the ART vigilance parameter without altering the overall neural structure. In the first experiment, the ART modules representing cortex used a vigilance of 0.8. For the second experiment, we increased vigilance in each of the ART modules within the first and second levels of cortex to 0.9 and 0.85 respectively (Figure 6). The first two levels of cortex in the model create the initial representations of the input images.

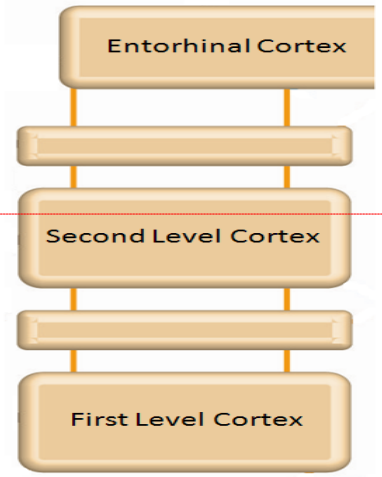

Figure 6: Architecture Sub-structure

\Increasing the precision of these initial representations allows the model to distinguish between similar inputs and additionally produces a more crisp recall. The second experiment atones to this improvement as shown by the results given in Table 2. All of the houses are captured in this experiment, and only Face4 yielded an incorrect association. We believe the reason for this incorrect recall when cued with Face4 is due to the striking similarity between faces one and four. Along this line of reasoning, Face4 errantly recalls House1, which is the house correctly associated with Face1. Besides the binary assessment of correct or incorrect recollection, the improved performance of the model in the second experiment is also apparent by the magnitude of the NCD values. The most specific recalled image from our first experiment yields a NCD value of 0.23, as compared with the least specific recalled image in our second experiment which has a NCD value of 0.27. Additionally, the very small NCD values in the second experiment, less than 0.1, indicate a very close match to the original house images.

	House1	House2	House3	House4
Face1	0.0087	0.8523	0.8538	0.8432
Face2	0.8471	0.094	0.8596	0.8442
Face3	0.8519	0.861	0.2701	0.8464
Face4	0.0087	0.8523	0.8538	0.8432
Face5	0.0087	0.8523	0.8538	0.8432
Face6	0.8471	0.094	0.8596	0.8442
Face7	0.8519	0.861	0.2701	0.8464
Face8	0.849	0.8571	0.8594	0.1054

Correct Answer | Incorrect Answer

Table 2: Increased Vigilance Quantitative Results

7. CONCLUSIONS

Qualitative analysis provides a means of assessing whether the model performs at a functional level comparable with the neuroanatomy that the architecture design was based upon. Our quantitative analysis provides a more rigorous means of analyzing models. The NCD distance metric is not tied to any particular features of the model, but rather is a mathematically rigorous universal distance metric applicable to other neural modeling problems as well. Specifically, it serves as a mechanism working towards the ability to both quantitatively assess neurocomputational models as well as to compare various models to one another despite differences in implementation details and other limiting factors. Thus, in this paper we have presented a neurologically plausible artificial neural network computational architecture of episodic memory and recall modeled after cortical-hippocampal structure and function, with a information theoretic based quantitative mathematical assessment.

8. REFERENCES

[1] Carpenter, G.A. & Grossberg, S. (2003), Adaptive Resonance Theory, In Michael A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, Second Edition (pp. 87-90). Cambridge, MA: MIT Press

[2] Cilibrasi, R., Vitanyi, M. Clustering by Compression. *IEEE Transactions on Information Theory,* 1523-1545.

[3] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vit´anyi, "The similarity metric," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3250–3264, 2004.

[4] Preston, Alison R. "Hippocampal Contribution to the Novel Use of Relational Information in Declarative Memory." Hippocampus (2004): 148-52.

[5] Quach, Tu-Thach. "Information similarity metrics in information security and forensics." Diss. U of New Mexico, 2009.

[6] Scoville, William B., and Brenda Milner. "Loss of Recent Memory After Bilateral Hippocampal Lesions." J Neuropsychiatry Clin Neurosci 12 (2000): 103-13.

[7] Turing, Alan M. "Computing Machinery and Intelligence." *Mind* LIX (1950): 433-460.

[8] Vineyard, Taylor, Bernard, Verzi, Morrow, Watson, Eichenbaum, Healy, Caudell, and Cohen. Episodic Memory Modeled by an Integrated Cortical-Hippocampal Neural Architecture. Human Behavior and Computational Modeling Conference 2009, June 23-24 2009, Oak Ridge TN, United States