

# Long-term Memory Structure by Hierarchized SOM

Paulo Salgado  
CITAB/ECT Depart. De Engenharias  
Universidade de Trás-os-Montes e Alto  
Douro  
Vila Real, Portugal  
psal@utad.pt

Getúlio Igrejas  
Escola Superior de Tecnologia e Gestão  
Instituto Politécnico de Bragança  
Bragança, Portugal  
igrejas@ipb.pt

Paulo Afonso  
Instituto de Telecomunicações  
ESTGA, Universidade de Aveiro  
Aguada, Portugal  
pafnaa@ua.pt

**Abstract**—The most evolved living organisms are characterized by having a sophisticated central nervous system with memorizing capabilities. In its structure are the neuronal cells. Cognitive memory allows the recording, storage, renewal and reading of information. In this work, memory is characterized by its functional and structural aspects and the main mechanisms for the learning and storage processes are also presented. A hierarchical model based on artificial neural networks is proposed as a conceptual model of information transfer between short and long-term memory. Inspired in recent studies, this model is based on an hierarchized structure, composed by layers of neurons with an appropriated topology and learning mechanism suitable for the retention of information in an organized way. A set of simulations is performed to test and evaluate the performance of this neuronal structure, particularly on how the information is distributed.

**Keywords**—*Hierarchical memory, Self-organizing map, PCA, clustering*

## I. INTRODUCTION

The human brain is the most complex and extraordinary structure ever known. This organ is capable, among many other functionalities, of receiving information from the sensory organs of the peripheral nervous system (hearing, sight, smell, etc.), analysing it, processing it, and generating output stimuli. If necessary, it can store this information for future use [1]. It is imperative for the achievement of cognitive intelligence [2], participating in pattern recognition tasks and facilitating the organization of relevant information, making it available in different ways (speech, writing, movement, among others). All of these processes involve cellular, electrical and chemical neuronal changes in the central nervous system, which are carried out dynamically and at different temporal scales [3][4]. When subjected to new excitation signals, these structures may be strengthened, weakened or changed, by creating new synapses [5].

Recent studies revealed in more detail the structure of the human brain [6], which contains billions of neurons, many of them connected to ten thousand other neurons, by synapses, that together form a neural network. The mechanisms of memorization at the level of neuronal cells also begin to be understood [7]. But the way as the brain processes information, compiles and distributes it through the neural network of memory is still poorly understood, namely the connection between its short and long memory.

One of the most recent theories of the organization of long-term memory (LTM) is Connectionism. Its principle is that mental phenomena can be described by interconnected networks of elementary units of processing (neurons) strongly

interconnected (by synapses), within prevailing the approach of a parallel distributed processing mechanism, spreaded through the networks [8][9]. In this structure, units or nodes represent a concept while connections between them represent learned associations. Without fault of generalization principle, in this paper these concepts are materialized by the weight vector of units, contextualized by its localization in the network. The learning mechanism explores the network interconnections and the adjustment of weighted units to expose the concepts, taking in account that the activation of a node will activate other nodes associated with it, which could work together to process a single memory. Although this model does not need necessarily to be biologically realistic, it has interpretable similarities and behaviour [10].

This study aims to show the structure of the brain related to the cognitive processes of memorization. Moreover, this approach is supported by recent evidences that suggest the memory is organized in a hierarchical way [11][12].

Its main objective is to propose a memory structure: Hierarchized Memory Structure (HMS), which represents a conceptual model of the memory processes. This is partly motivated by how visual, auditory or other sensory information is diffused and retained throughout the separated parts of the cerebral cortex in the human brain [8]. Its structure comprises two levels. The first consists in a hierarchical structure organized into layers of neurons and inter-layered connections. The second level is formed by the arrangement of neurons and intra-layer connections, which obey a semantic typology, with some similarities to Self-Organizing-Maps (SOM) [13] [14]. The goal of unsupervised learning in the HMS is to cause different parts of the network to respond similarly to certain input patterns, which is a very useful tool in discovering knowledge from data. However, this activation map is now a three-dimensional visualization for the resulting groups.

This model was tested in some examples. The results show that HMS is able to recognize and memorize information in the form of patterns while organizing and distributing it through different layers that retain certain characteristics of the data.

This paper is organized as follows. In the following section, cognitive memory of human brain is reviewed and its main features presented. In Section 3 the HMS is presented and its hierarchical structure and topology are analysed. Next, a learning algorithm for the HMS memory model is proposed. The results of the computational tests of the HMS network are presented in section 5, where they are analysed and discussed. Finally, in the last section the main conclusions are presented.

## II. MEMORY

Memories are stored distributed throughout the brain [15]. When it performs a remembering task, the brain is able to gather all the relevant stored fragments, even if they are scattered in the most diverse places of the brain, uniting them and giving it meaning. The result is usually complete and organized information. However, these mechanisms are still badly understood [16][17].

Memories can reside in the short-term "buffer" for a limited time [18]. These type of memories are processed in the front of the brain, in a highly developed region called the prefrontal lobe. If needed they can be converted into LTM by the hippocampus and migrate to deeper areas of the brain. The LTM have two major headings: explicit (or declarative) memory and implicit (or procedural) memory [19]. The first one is encoded by the hippocampus, entorhinal cortex, and perirhinal cortex, but consolidated and stored elsewhere in a not fully defined region of the medial temporal lobe of the brain [20]. The second is encoded and it is presumably stored by the striatum and other parts of the basal ganglia [21].

In this work we are focused in LTM by combination of two headings: hierarchic and semantic networks in the same structure. This contains interconnected nodes and each node represents a concept. These conceptual nodes are connected or linked according to their relationship. Concepts may represent physical objects, events, attributes, or abstractions. In the hierarchical perspective, they are usually arranged from general to more specific classes or from simple to complex. Otherwise, in the semantic perspective, memory is now organized into loosely connected concepts where certain triggers activate associated memories according to their relationship. Although it has similarities to hierarchies, semantic networks are more random and less structured than true hierarchies. They have multiple links from one concept to others. Concepts within semantic networks are not limited to specific aspects, but perhaps based on the meaning and relationships that you have learned through experiences. New memories are formed by adding new nodes to the network and information needs to be linked to existing networks memory. Therefore, new information is placed in the network by connecting it to the appropriate nodes. If information is not associated with existing information it is forgotten.

## III. THE HIERARCHIZED MEMORY STRUCTURE

The LTM encodes information semantically for storage through a process referred as Synaptic Consolidation, where items are transferred from short-term memory (STM) to LTM [22]. Memory traces in the cortex are embedded into a feed-forward and recurrent connectivity that is a functional hierarchical structure. The consolidation of information into memory network depends on repetitive and off-line processes. It is recognized the role of sleep state or during restful waking brain activity in this task [23]. However, free to not present here an unequivocal answer about these subjects, it will be presented in this work a memory based conceptual solution using an hierarchized structure with knowledge layers. For this structure an algorithm, which rules determine how information is distributed along the network memory in an unsupervised manner, is proposed. For reasons of simplicity of analysis, this

network will be not implemented with feedback-learning strategy.

### A. Hierarchized Memory Structure

The HMS is an unsupervised neural network that maps a set of  $n$ -dimensional vectors (input vector  $x$ ) into to a three-dimensional topographic map structured with  $L$  layers, each one with nodes arranged into a regular grid in a finite two-dimensional region. Each node (or neuron)  $i$  of  $k$ th layer is associated with a "weight" vector,  $W_{ik}$ , with the same dimension of the input vector  $x$ , that is a position in the input space,  $P_{ik}$ , generally fixed, ensuring the topological shape of the map space. The number of connected neurons inside the layer is finite and determined by the maximum influence radius of intra-neurons,  $r_I$ , and the metric space considered (for example, plane, cylindrical or spherical). Likewise, the inter-neuron radius of influence,  $r_L$ , determines the connections between consecutive neurons layers.

Fig. 1 shows a three layers HMS model with a hexagonal topology and a grid of  $[10 \times 6]$  and  $[3 \times 3]$  nodes, respectively, for the second and third layer.

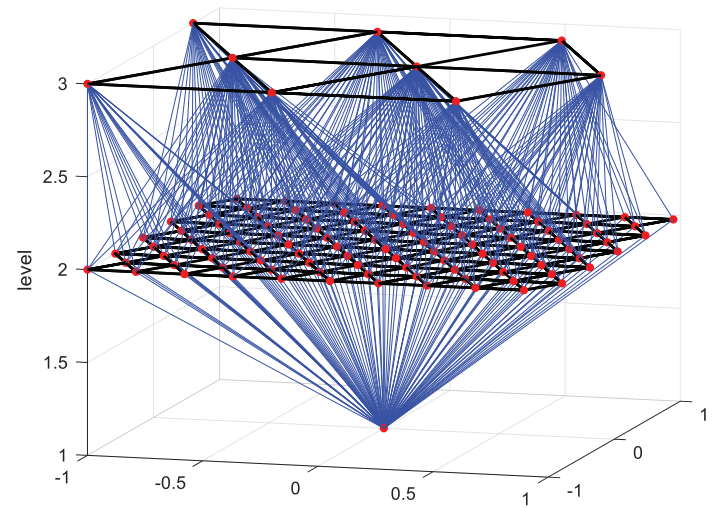


Fig. 1. Topology of HMS model, with  $L=3$  layers.

During the memorization process the weight vectors are moving towards the input data, while  $P_{ik}$ 's are static. Due to the dynamics of inherently competitive unity in each layer, the winner takes everything to assimilate as much information as possible. The remaining part of information transits to the next layers of the hierarchical structure to be assimilated. This way, the information is distributed over the network, each layer retaining a directional component of the information. The kind of information retained by each layer depends on the learning method and similarity measure used in that layer.

### B. Learning Algorithm

In the HMS, the goal of learning is to cause different parts of the network, one for each layer, to respond similarly to certain input patterns. In the first step, input  $x$  is presented to second layer of the network, after going through the first layer with (one) transparent(s) neuron(s). Each data, randomly

chosen, is presented, one by one, to the input HMS until the data train set is exhausted. So, at discrete time  $t$ ,  $x(t)$ , with  $t = 0, 1, \dots$ , is presented to the network and all weight vector,  $w_k(t)$  are tuned, for  $k=2, \dots, L$  layers. The winner node  $s_k$  of the  $k$ th layer is the one which has the closest weight vector,  $W_{s_k}$ , to the input vector  $X_{s_k}$ . For the second layer,  $k=2$ , this similarity is determined using Euclidean distance, i.e.

$s_k = \min_i \|x(t) - W_{i,k}(t)\|$ , while for the upper layer this is done by using the maximum cosine similarity value,  $s_k = \max_i (\phi_{ik}(t) \cdot |W_{ik}^T(t)x_{k-1}(t)|)$ , where  $x_{k-1}$  is the output of the

winner neuron of the previous layer and  $\phi_{ik}(t)$  is the weight inter-layers of the neighbourhood function. Usually, the initial values of the  $w_k$  are chosen from the data set or by tanking a small random vector value. Their weights and of its surrounding nodes are adapted recursively using equation (1):

$$W_{i,k}(t+1) = W_{i,k}(t) + \alpha(t) \cdot \varphi_{i,k}(t) [x_k(t) - W_{i,k}(t)] \quad (1)$$

In equation (1),  $i \in N(s_k)$  for the second layers ( $k=2$ ) and  $t$  is the current time,  $\alpha(t)$  is the learning rate and  $\varphi_{i,k}(t)$  is the intra-layer neighbourhood factor of the  $i$ th neuron of the winning neuron,  $s_k$ .

For the next layers, the inputs are the remaining part of the network input that was not “memorized” by the winner neurons of the previous layers. So, the input of the  $i$ th neuron of  $k+1$  layer is given, for  $k=2$ , by equation (2):

$$x_{i,k+1}(t) = \phi_{is,k}(t) [x_k(t) - W_{s_k}(t)] \quad (2)$$

or, for  $k>2$ , by equation (3).

$$x_{i,k+1}(t) = \phi_{is,k}(t) [I - W_{s_k}^T(t)W_{s_k}(t)] x_{s_k}(t) \quad (3)$$

In equation (3)  $\phi_{is,k}(t)$  is the neighbourhood interlayer factor between the winning neuron,  $s_k$ , of the  $k$ th layer and the  $i$ th neuron of the next layer. In the second layer, the error values of the winner (the difference between the input vector and the weight vector) transits to the next layer. Otherwise, it is an orthogonal projection vector to the  $W_{s_k}(t)$ , with unit norm,  $\|W_{s_k}\| = 1$ .

The update of the weight vector is done, for  $k > 2$  and  $i \in N(s)$ , by equation (4).

$$W_{i,k}(t+1) = W_{i,k}(t) + \alpha(t) \cdot \varphi_{i,k}(t) (W_{ik}^T(t)x_{ik}(t)) \quad (4)$$

During the weight adaptation, the learning rate used is reduced along iterations according to the total number of current

nodes, from a value  $a$  to  $b$ , i.e.:  $\alpha(t) = a + (b-a)(1-\tilde{t})$ , with  $\tilde{t} = t/t_{\max}$ .

In this work the interlayer and interlayer neighbourhood are a Gaussian functions, represented by equation (5) and equation (6), respectively.

$$\varphi_{ik}(t) = \exp\left(-\frac{\|P_{s_k} - P_{ik}\|^2}{(2R_k^2)} + \tilde{t} \cdot M \cdot \neg L_k(s,i)\right) \quad (5)$$

$$\phi_{ik}(t) = \exp\left(-\frac{\|P_{s_{k-1}} - P_{ik}\|^2}{(2R_k^2)} + \tilde{t} \cdot M \cdot \neg L_{k-1,k}(s,i)\right) \quad (6)$$

In equation (5) and (6),  $L_k(s,i)$  is the link weight intra-layer, between the neurons  $s$  and  $i$  of the  $k^{\text{th}}$  layer, and  $L_{k-1,k}(s,i)$  is the link weight inter layers, between the  $s^{\text{th}}$  neuron of  $k-1$  layer to the  $i^{\text{th}}$  neurons of layer  $k$ , which are logic values.  $M \rightarrow \infty$  is a parameter of high value that penalizes no linked neurons.  $R_k(t)$  is the neighbourhood function that decreases during train, from an initial value comparable to the dimension of the lattice,  $R_{mk}$ , to a final value effectively equal to the width of a single cell,  $R_{0k}$ :

$$R_k(n) = R_{mk} + \frac{t}{t_{\max}} (R_{mk} - R_{0k}) \quad (7)$$

This process is iteratively repeated for a sufficient number of cycles, using all the data set. For each layer, the network ends up associating their nodes with groups or patterns in the input layer space, storing the maximum selective information possible. The selectivity propriety of each layer is dependent on the number of neurons, but essentially on topology, links intra and inter layers and similarity measures used in each layer. The remaining information of the present layers will be used as input for the next layer until arriving to the last layer or non-information exists to memorize. At the end of this learning process, weight vectors are prototypes representing the input data in the layer context. This process is repeated until the output map converges to a stable or organized state, when the average error falls below a pre-specified value or a certain number of iterations have been reached.

#### IV. RESULTS

The HSM was tested with two examples. The first one aims to memorize a set of data shown in Fig. 2. It is composed by 4 ellipsoidal clusters with angular orientation about  $\pm\pi/4$  and centres  $(\pm 0.5, \pm 0.5)$ . Each cluster has 200 elements in  $\mathbb{R}^2$ . The main objective of the second layer of HMS is to identify the clusters centres while the last layer finds their directions. The data set was used, one by one element, repeatedly into 20 iterations to tune the  $W$ 's parameters. At the end of this learning process, the positions of the weights memorize information's

features of data set. Fig. 3 shows the positions of  $W$ 's vectors in the Euclidean space at the end of the memorizing process, layer by layer.

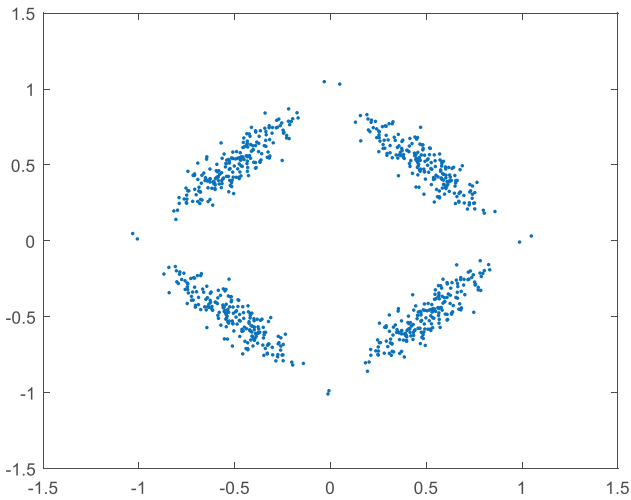


Fig. 2. Training data (4 clusters with orientation  $\pi/4 \pm j\pi$ ).

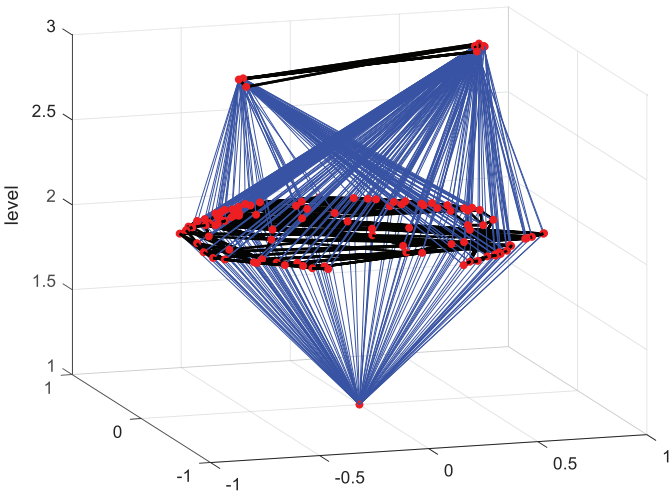


Fig. 3. U matrix representation of HMS (layer by layer) with inter and intra-connections (blue lines).

We conclude that weights of second layer of HMS represents the centers of clusters of data (located around the following 4 positions:  $(\pm 0.5, \pm 0.5)$ ). Even more, the weight vectors of 3<sup>rd</sup> layer, with values  $[-1; -1]$  and  $[1; 1]$ , are the principal component of data cluster distribution, with orientation values around  $\pm \pi/4$  rads. So, the first layer memorizes centres of clusters while the 2<sup>nd</sup> layer stores its orientation.

In the second test example, the HMS has a 4<sup>th</sup> layer structure, with a configuration topology of  $4 \times 4$  neurons, for the  $k=2, 3$  and 4, as shown in fig. 5. The data set used has a distribution on a two-dimensional space with a triangular shape (see Fig. 4). Their edges are dispersive data points according three directions.

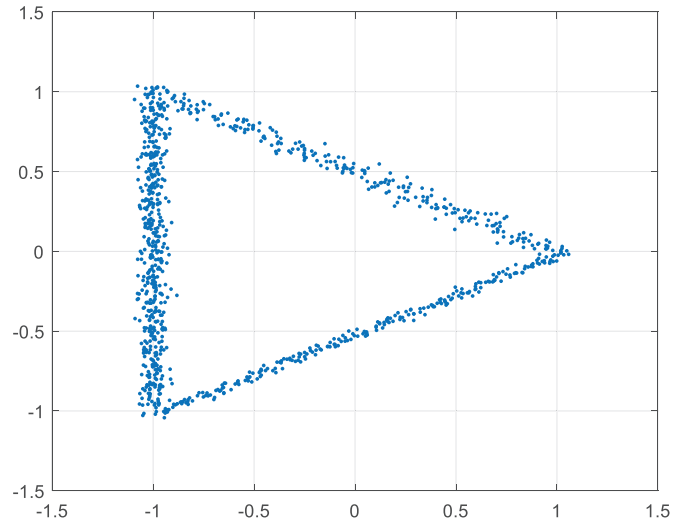


Fig. 4. Data to be memorized by HMS (data draw a trinagule with 3 edges).

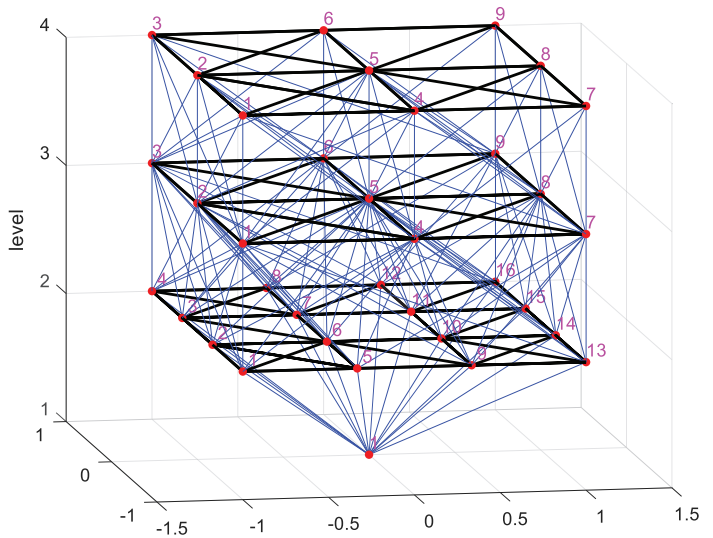


Fig. 5. Topology of HMS structure.

After one hundred iteration cycles of train, the weight of neurons in each layer migrates to capture the most possible information according to the similarity used. For the 2nd layer the Euclidean distance was used while for the 3rd and 4th layers we used the cosine similarity measure. The results of this process is presented in Fig. 6. Fig. 7 shows the neuron activity for the data set (with one thousand points).

From Fig. 7 is possible to observe that neurons 6 and 10 of the 2nd layer are not used (and so available to be used for memorize other kinds of information). Weights of the 2nd and 3rd layers are, respectively, the principal component vectors of the distributed data around the neighbourhood region of neurons of 2nd layer.



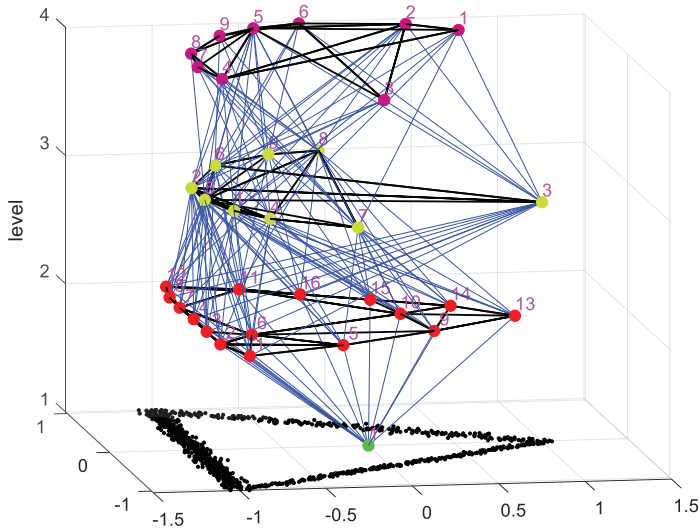


Fig. 6. U matrix representation of neuros weight, for the 4 layers.

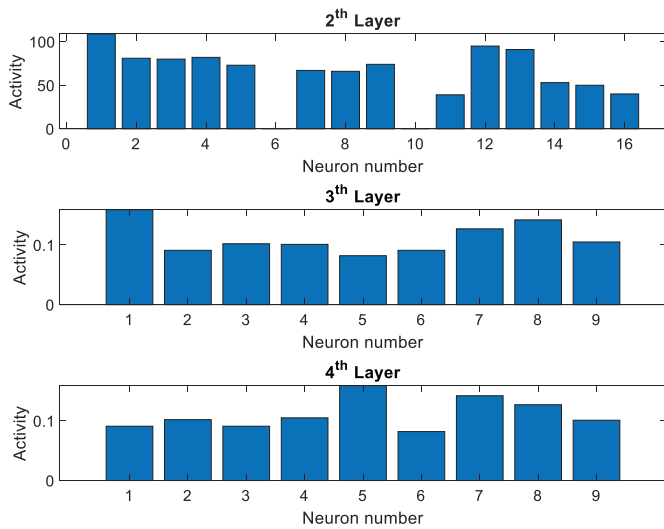


Fig. 7. Activity of neuros for each layer (for the thousand data size of test example).

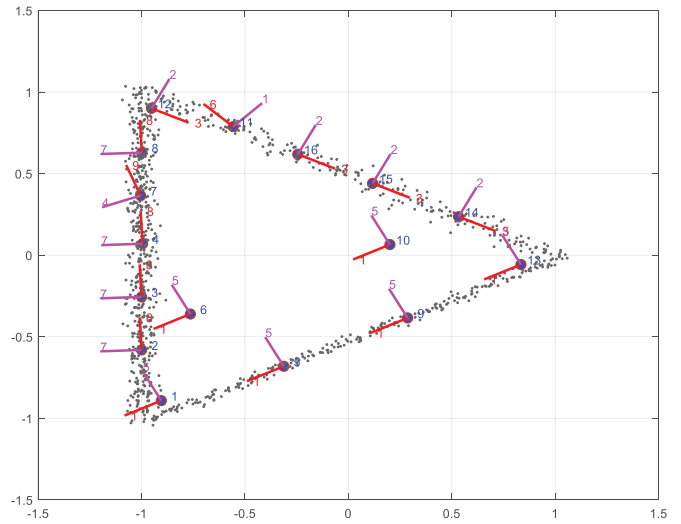


Fig. 8. Weights of 2nd Layer (points) and of 3rd (vectors).

In Fig. 8 is aggregated in one figure the data distribution (to memorize), the weights of the 1st layer (by blue circles) and the vectors (shorted to magnitude 0.2) of weight vector of the 2nd layer (red vector) and 3rd layer, respectively, by red and magenta line colours. Vectors represent the weights of last layer while the points represents the weights of previously layer. As observable, these vectors are pointing in the direction of the maximum variation of data or in its orthogonal direction.

Taking the results of these two examples, HMS may be considered a nonlinear generalization of Principal Component Analysis (PCA). The second layers found the centers of clusters while the upper layers done successively the principal, secondary, and so on, component vectors of the clusters correlation matrix. During the training phase, weights of the whole neighborhood are moved in the same direction, similar items tend to excite adjacent neurons. So, HMS forms a semantic hierarchized map where similar features of samples are mapped close together and dissimilar ones apart. As was shown in fig. 8, W's are discrete approximation of the distribution of training samples.

TABLE I. ANGLES OF WEIGHTS VECTORS OF THE 2SD AND 3RD LAYERS AND ITS ACTIVITY (VALUES INSIDE PARENTHESIS).

Layer	Neuron n°								
	1	2	3	4	5	6	7	8	9
3	27.2 (38%)	-8.3 (0%)	-26.1 (25%)	42.8 (0%)	9.3 (6%)	-44.0 (6%)	72.9 (0%)	-87.4 (18%)	69.5 (6%)
4	45.8 (6%)	64.2 (25%)	81.4 (6%)	20.9 (6%)	-63.3 (31%)	-80.7 (0%)	2.5 (25%)	-17.4 (0%)	-46.9 (0%)

## V. CONCLUSIONS

Firstly, this paper summarizes the actual knowledge about the brain memory process, at the level of cell neurons and at the field of neurology . Two main memory types are identified: the STM and LTM. Their main mechanisms and characteristics were briefly described. Based on these results and hypothesis about how the memory is processed in the brain, it was proposed

an Hierarchical Memory System with a learning algorithm to store information into a distributed and stratified layers structure. This model was validated by two example tests. In both, it shown good performance to memorize information contained into numerical vector data set. Our experiments demonstrated the basic effectiveness and adaptability of the HMS into the clustering task and in memorizing objects into a distributive and stratified hierarchical structure. The HMS supported by appropriate topology, distance and similarity measures can memorize properly, with layers to reveal features of data or of information. Thus, the HMS appears to be a useful technique to decompose and consolidate information, but also to memorize it into an organized way. It can also to be used to cluster or recognize objects.

## REFERENCES

- [1] Information Storage, Learning and Memory. In: Stolerman I.P. (eds) *Encyclopedia of Psychopharmacology*. Springer, Berlin, Heidelberg, 2010
- [2] Bontempi Bruno Silva Alcino J. Christen Yves, *Memories: Molecules and Circuits*, Springer-Verlag Berlin Heidelberg, 2007, ISBN: 978-3-540-45698-8
- [3] Robert E. Clark Stephen J. Martin, *Behavioral Neuroscience of Learning and Memory*, Springer International Publishing, 2018, ISBN: 978-3-319-78755-8
- [4] Neurotransmitters Used by the Axons that Form Synapses Onto These Neurons. In: Gebhart G.F., Schmidt R.F. (eds) *Encyclopedia of Pain*. Springer, Berlin, Heidelberg, 2013
- [5] L. Andrew Coward, *Towards a Theoretical Neuroscience: from Cell Chemistry to Cognition*, Springer Science+Business Media Dordrecht 2013, ISBN: 978-94-007-7106-2
- [6] Brain Mapping. In: Stolerman I.P. (eds) *Encyclopedia of Psychopharmacology*. Springer, Berlin, Heidelberg, 2010.
- [7] Burger J.R., *Brain Structure*. In: *Brain Theory From A Circuits And Systems Perspective*. Springer Series in Cognitive and Neural Systems, vol 6. Springer, New York, NY, 2013.
- [8] Jerry A.Fodor and Zenon W.Pylyshyn, *Connectionism and cognitive architecture: A critical analysis*, *Cognition*, vol. 28, Issues 1–2, March 1988, Pages 3-71.
- [9] A.Ellery , *Artificial intelligence through symbolic connectionism—A biomimetic rapprochement*, *Biomimetic Technologies - Principles and Applications*, Woodhead Publishing Series in Electronic and Optical Materials, 2015, pp. 227-252, <https://doi.org/10.1016/B978-0-08-100249-0.00011-2>
- [10] Stephen Grossberg, *From brain synapses to systems for learning and memory: Object recognition, spatial navigation, timed conditioning, and movement control*, *Brain Research*, Vol. 1621, 2015, pp. 270-293, ISSN 0006-8993, <https://doi.org/10.1016/j.brainres.2014.11.018>. (<http://www.sciencedirect.com/science/article/pii/S0006899314015650>)
- [11] Mortimer Mishkin, W. Suzuki, D. Gadian and F. Vargha-Khadem, *Hierarchical organization of cognitive memory*, *Phil.Trans. Real Society London B* , Vol. 352, pp. 1461-1467, 1997.
- [12] Tomoki Kurikawa and Kunihiro Kaneko, *Dynamic Organization of Hierarchical Memories*, *PLOS ONE* DOI:10.1371/journal.pone.0162640 September 12, 2016
- [13] Kohonen T, *The self-organizing map*. *Proc IEEE* 78:1464–1480, 1990
- [14] Yin, Hujun, *The Self-Organizing Maps: Background, Theories, Extensions and Applications*, Ed.Fulcher, John and Jain, L. C., *Computational Intelligence: A Compendium*, 2008, Springer Berlin Heidelberg, 715-762, ISBN 978-3-540-78293-3, doi="10.1007/978-3-540-78293-3\_17",
- [15] Brain Mapping. In: Stolerman I.P. (eds) *Encyclopedia of Psychopharmacology*. Springer, Berlin, Heidelberg, 2010
- [16] Juri D.Kropotov, *Memory Systems, Functional Neuromarkers for Psychiatry, Applications for Diagnosis and Treatment*, 2016, (Chapter 3.4), pp. 231-242, <https://doi.org/10.1016/B978-0-12-410513-3.00014-0>
- [17] Izquierdo I., Myskiw J.C., Benetti F., Furini C., *Protein Synthesis and Memory*. In: Stolerman I.P., Price L.H. (eds) *Encyclopedia of Psychopharmacology*. Springer, Berlin, Heidelberg, 2015.
- [18] Cools R., *Short-Term and Working Memory in Humans*. In: Stolerman I., Price L. (eds) *Encyclopedia of Psychopharmacology*. Springer, Berlin, Heidelberg, 2014.
- [19] Mark W. Schurkin, *Visual memory, the long and the short of it: A review of visual working memory and long-term memory*, *Attention, Perception, & Psychophysics*, July 2018, vol. 80, Issue 5, pp. 1035–1056.
- [20] Moscovitch M. et al., *Hippocampal Complex Contribution to Retention and Retrieval of Recent and Remote Episodic and Semantic Memories: Evidence from Behavioral and Neuroimaging Studies of Healthy and Brain-Damaged People*. In: Ohta N., MacLeod C.M., Uttl B. (eds) *Dynamic Cognitive Processes*. Springer, Tokyo, 2005.
- [21] J. Goodman, M. G. Packard , Chapter 35: *Memory Systems of the Basal Ganglia*, *Handbook of Behavioral Neuroscience*, vol. 24, 2017, pp. 725-740.
- [22] Yadin Dudai, Avi Karni and Jan Born, *The Consolidation and Transformation of Memory*, *Neuron*, vol. 88, Issue 1, 7 October 2015, pp. 20-32, <https://doi.org/10.1016/j.neuron.2015.09.004>.
- [23] Monika Schönerauerabc, Melanie Grätschb and Steffen Gais, *Evidence for two distinct sleep-related long-term memory consolidation processes*, *Cortex*, volume 63, February 2015, pp. 68-78, <https://doi.org/10.1016/j.cortex.2014.08.005>.