Attractor Neural Network Approaches in the Memory Modeling

Berat Denizdurduran

Electronics and Communications Engineering Department,
Istanbul Technical University,
34469, Istanbul, Turkey
denizdurdu@itu.edu.tr

Abstract. Memory is an ability to store the experiences to understand the new environmental conditions. Recent physiological experimental results clarify the hippocampus role in memory and spatial navigation. There are different approaches to modelling the memory and spatial navigation, such as neural networks and dynamical systems. The main point of this models are to show the efficiency of the brain inspired computational models, also explain the connections with biological details.

Keywords: Hippocampus, memory, spatial navigation, cognitive map, dynamical systems, computational models.

1 Introduction

In order to reach goals, agents need some behavioural and cognitive skills such as action selection, learning, memory, spatial navigation and environmental orientation. Memory role in a complex task can be defined as the process of the retrieving. It allows us to compare the experiments. The interest on the definition of memory has been started long before the contemporary neuroscience. The first step of the theory might be started with Aristotle's ideas. According to Aristotle, knowledge of the objects depend on feelings without their substance like a ring's trace in the wax [1]. Aristotle has seen the memory as trace of sensations. Plato expanded the idea of Aristotle with include the memory storage phenomena. According to the Plato, the trace of sensations could be stored but it might be mistakes during the recall process [2].

One of the most challenging questions in contemporary neuroscience is what is the secret of the brain. Where the knowledge store and how is it encode or discriminate? Which substructures are responsible for these processes? There are several studies, which strongly claim that hippocampus has a vital role in memory and cognitive map [3]. The brain-damaged patients have major contributions to many scientific works, even perhaps reluctantly. These patients allow the scientists to compare their theory of related brain regions. The first observations of the hippocampal functions also depend on brain-damaged patient study which belong to William Scoville and Brenda Miller in 1957. Furthermore, they used some special technology to screen the corresponding neurons behavior. The results show the significant correlation between cellular pattern with sensory and behavioural variation [4]. O'Keefe et.al. suggested the cognitive map theory which based on these experimental results, also the spatial functions of the hippocampus indicated in this work [5]. Even the general idea of the hippocampus, which is thought like a memory-related brain region, is well-known, there are still being some lack of knowledge to the other potential functions. For example, there is a link between hippocampus and hypothalamus and it is assumed to relate on hippocampus' role in stress responses [6].

The brain inspired computational models are exploring the link between neurons and behaviour. First, the models aim to purpose a hypothetical mechanism about the knowledge of brain regions and their connections. Then, these mechanisms enable the scientists' ideas to constitute the precise and quantitative theories. To understand the hippocampus functions, for more than forty years, there are several computational models exist. One of the approach to model the brain regions is dynamical systems [7]-[12]. In the beginning of contemporary neuroscience, David Marr suggested a pioneered theory to describe the episodic memory [7]. Follow from Marr's model, O'Keefe et.al. considered the memory as attractor states to represent the neuronal networks [8]. In neuroscience literature, the spiking neural network models are commonly used to build the structure of the networks beside dynamical systems. In [9], the integrate-and-fire neuron model is used to model the hippocampal place cells. The main point of this work is to show the relationship between the phase response and location. In the well-known work of cognitive maps [10], some specific neurons, which are known as place neurons, are constructed with attractor set. Then, attractor map is occurred by these attractor set. This model represents the relationship between the graph and sensory inputs to explain the coordinates of arbitrary environments. Pioneering works on memory are explained the relation between associative memory and the CA3 region of the hippocampus [11],[12]. When think all these computational models together, it can be suggested that knowledge of the human brain and the idea of hippocampal formation on memory are well-known day by day.

2 Attractor Neural Networks

Neuron's behaviour can be described as all-or-none responses. The activation of the membrane potential depends on the current coming into the neuron. In a given network, the highest firing rate is important for the neuron's behaviour. In computational neuroscience, it is also known as winner-take-all states. The environmental conditions have a role to control the electrical discharge of neurons. In contrasts, the acts of neurons can be different regarding to the substructures of brain, timing or connections. In order to model the neuron behaviour, attractor states are thought to be suitable. The connections between neurons have a decisive role to stabilize the corresponding networks [13]. These connections are evaluated by using Hebbian Learning in neuronal modelling [14].

Briefly, in the Hebbian Learning, the input current effects the neuronal activity and changes the transfer function. This activation is given by:

$$\alpha = \omega \upsilon \ . \tag{1}$$

A connection weight ω denotes the modification of learning and υ denotes the sum of firing rates. The connection weights are evaluated by:

$$\tau \frac{\mathrm{d}\omega_{\mathrm{i}}}{\mathrm{d}x} = \alpha \upsilon_i \ . \tag{2}$$

where ω gives the rate of change of connection weights with time. During the training, total change in ω:

$$\omega \rightarrow \omega + \frac{T}{\omega} \sum_{\mu} \alpha^{\mu} \upsilon^{\mu} \ . \tag{3}$$

evaluation of ω after presentation of all input patterns:

$$\omega \rightarrow \omega + \frac{T}{\omega} \sum_{\mu} (w \upsilon^{\mu}) \upsilon^{\mu} \ . \tag{4}$$

Another call of Hebbian learning rules is correlation-based learning rules. In some cases, such as large weights during the evaluation, this learning rule might not be stable. In this case, output activation might not be in the significant range. The connections between two neurons depend on the timing and it might be strengthened when neurons are fired simultaneously. The unstable situation occurs where the stimulation time is pre- or post-connection activity. In Hebb synapses, connection strengths can grow without limits, as mentioned above. To deal with this issue, Caianiello suggested the Adiabatic Learning Hypothesis [15]. In this hypothesis, regarding to the system dynamics, the process of learning is particularly slow. The other perspective to solve this problem is synaptic adaptation model. In this work, learning and neuronal activity are considered in separate time scales [16].

The Hopfield network [17], which considers the Hebb's idea of synaptic plasticity, is the beginning of attractor neural networks model approaches. The traning of patterns in memory:

$$\omega_{ij} = \sum_{a} \xi_i^p \xi_j^p . \tag{5}$$

where $\xi_{i,j}$ ϵ (-1,1). The update state is evaluated by:

$$x_i \leftarrow sign(\sum_{j} \omega_{ij} x_j). \tag{6}$$

In Hopfield model, to define the state function, a particular function is used (a.k.a energy or lyapunov function). This function range is limited [16]. Attractor neural networks can be interpreted as a working memory where the weight matrix acts as a long-term memory [18]. Wills et.al suggested that memories are acting like attractor states. They recorded the activity of hippocampal place cells from rats in foraging task. The environmental shape is different in each task, such as square, circular or semisquare-semicircular [8]. To demonstrate the attractor states, inspired from [13] and [19], following figure is given. As it is shown in Figure 1a, during the rats' exploration in the environment, each place is denoted by a different attractor state. Different demonstrations represent the place neurons of the hippocampus. In attractor neural network approaches, a new attractor state may develop in order to recognize the new environment which depends on experiences (Figure 1b).

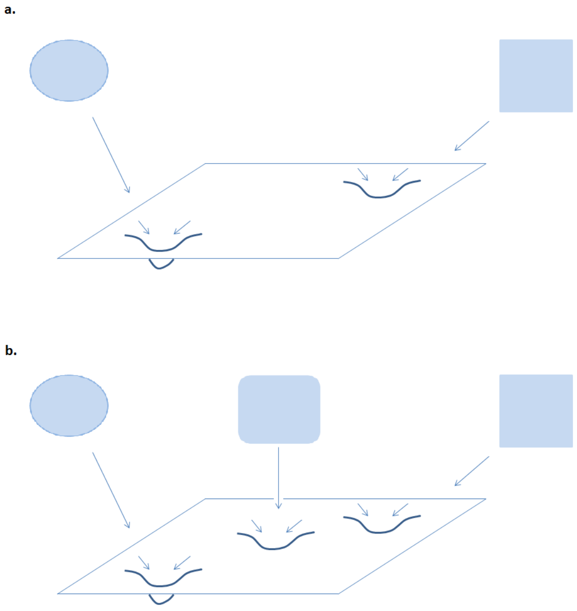

Fig. 1. Illustration of attractor states which demonstrate the different environments.

2.1 Point Attractors

In an attractor network, neurons are recurrently connected with each other with a finite number of connections. A stable states of the system can be described as attractor which is the result of corresponding connected neurons. This process allows us to describe the patterns under the dynamics of network. Each attractor states stored a patterns in the network.

In the process of retrieving this stored patterns, system needs a similar enough pattern activation then converged pattern is retrieved from the dynamics of network [20]. Such networks are also known as autoassociative networks which is an example of content-addressable memory. The simplest model of content addressable memory is Hopfield model [16] where connection weights evaluate form the Hebbian learning rules [14], as mentioned above. A biologically realistic model of autoassociative network is proposed in [21]. This work explains the relationship between CA3 region and autoassociator approach to explain the declarative memory. The synaptic modification is the main process of this idea. This process allows the storage and retrieval phenomena in the CA3 region of the hippocampus. In order to explain the structure of point attractors, a dynamical system's phase space is given in Figure 2. As it can be followed from the figure, at corresponding initial values, each attractor state has a different trajectory and each attractor can be described as convergence states of the system.

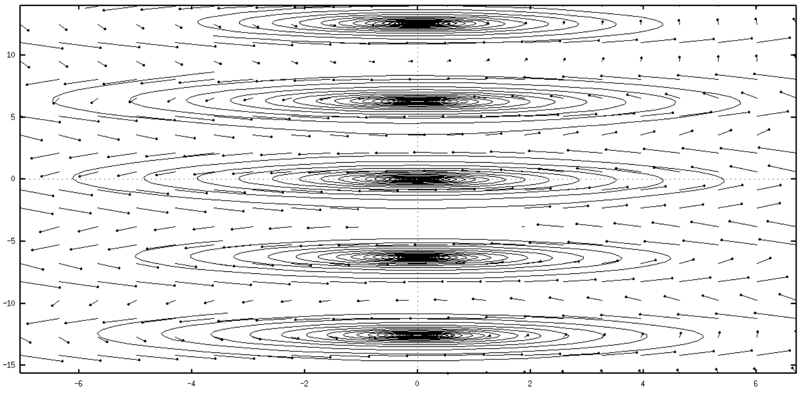

Fig. 2. Phase space of a dynamical system with local stability.

2.2 Line Attractors

The extension of point attractors can be defined with line attractors. There is an infinite set of points to state the fixed point (Figure 3). As it can be followed from the figure, initial values is not important for the system's behaviour but in some cases there are more than one line attractors in the system and region of the line attractors become important for the system's behaviour.

In [22], tuning curves describe the direction of the motion. Each pattern, which demonstrates the related neuronal stimulus in attractor states, indicates the corresponding location. A demonstration of the bell-shaped tuning curves is given, in order to represent the situation (Figure 4). The stimulus is shown like a smooth bump. The bump makes a peak connected with locational changes. These curves can be described as a function of the neurons activity and this activation depends on the current input during the exploration [23].

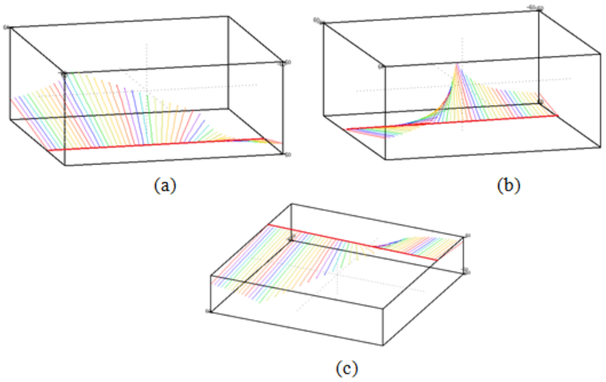

Fig. 3. Demonstration of line attractor. There are several trajectories which start in different initial values. Each trajectories is indicated with different colors. Red line indicates the attractor manifold. All of the subfigures belong to the same line attractor but illustrated in different angles.

Furthermore, the head direction cells are considered to describe the path integration in the environment [24],[25]. Regarding to the environmental changes, model show the adaptation of the systems in a given direction.

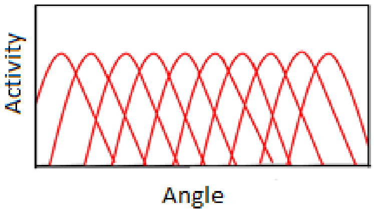

Fig. 4. Tuning curves describe the direction of the objects.

2.3 Continuous Attractors

Hippocampus has two well-known role in brain. One of them is mempory and the other one is spatial orientation. To exemplfy the spatial orientation role of hippocampus, a head-direction cells model is constituted with continuous attractors [24]. A detailed model of place cell representation with continuous attractor approach can be found in [10]. In this work, each recurrent connections are evaluated before the experiments. A Gaussian function is used to determine the environmental changes. The component of a gaussian function is described the part of the environment. Each evaluation depends on a rat's position in the environment [10].

The place field directionality is modeled with a recurrent network in [26]. The main focus in this study is to describe the model of entorhinal cortex from the point of locational and directional activity. Kali and Dayan suggested that the acquaintance of the environment can be affected from neuromodularity systems. First, the model learns the representation of any environment then model is exposed to another environment. If the novelty of second environment pattern is sufficient enough, novelty-modulated learning is possible [26]. This work shows the remapping of unrelated place cell representation.

3 Conclusion

The main point of brain inspired computational models is to suggest a link between the properties of cells in the brain and animal behavior. A main issue of this study was to show several computational models of how the hippocampus stores, retrieves and discriminates the memory and controls the spatial navigation. The mechanism of hippocampal function is modeled with different approaches, as it is same for the other neuroscience studies. One of the approaches in the memory modeling is attractor neural networks. The attractor paradigm of neural computation has an ability to explain the memory both its theoretical and experimental aspects.

Acknowledgments. This work is a part of mentorship program in IBRO/UNESCO Cape Town School on Advanced Theoretical and Computational Neurosciences. I would like to thanks to Prof. Misha Tsodyks for his guidance and sharing his knowledge.

References

1. Aristotle.: On memory and Reminiscence. (350).
2. Plato.: Theaetetus. (360).
3. Gaffan, D.: Loss of recognition memory in rats with lesions of the fornix. Neuropsychology 10:327-341 (1974).
4. Hirano, T., Best, P., Olds, J.: Units during habituation, discrimination learning, and extinction. Electroencephalography and Clinical Neurophysiology. 28:127-135 (1970).
5. OKeefe, J., Dostrovsky, J.: The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. Brain Research, 34-171-175, (1971).
6. Alvares, L. D. O.,Engelke, D. S., Diehl, F., Scheffer-Teixeira, R., Haubrich, J., Cassini, L. D. F., Molina, V. A., Quillfeldt, J. A.: Stress response recruits the hippocampal endocannabinoid system for the modulation of fear memory. Learning and Memory, 17:202-209, (2010).
7. Marr, D.: Simple memory: a theory for archicortex. Philosophical Transactions of the Royal Society B: Biological Sciences, 262:2381, (1971).

8. Wills, T., Lever, C., Cacucci, F., Burgess, N., OKeefe, J.: Attractor dynamics in the hippocampal representation of the local environment. Science, 308:873-876, (2005).
9. Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J., McNaughton, B. L.: Population dynamics and theta rhythm phase precession of hippocampal place cell firing: A spikin neuron model. Hippocampus, 6:271-280, (1996).
10. Samsonovich, A., McNaughton, B. L.: Path integration and cognitive mapping in a continuous attractor neural network model. The journal of neuroscience, 17(15):5900-5920, (1997).
11. Amit, D.J.: Modelling brain function. New York: Cambridge University Press, (1989).
12. Rolls E. T.: Functions of neuronal networks in the hippocampus and neocortex in memory. In: Byrne, J.H., Berry, W.O., editors, Neural Models of Plasticity. New York: Academic Press. (1989).
13. Andersen, P., Morris, R., Amaral, D., Biss, T., O'Keefe, J.: The Hippocampus Book. Oxford University Press, (2007)
14. Hebb, D.O.: The organisation of behavior. New York: Wiley, (1949).
15. Caianiello, E.: A theory of neural networks. In Aleksander, I., editor, Neural Computing Architecture, MIT Press, (1989).
16. Tsodyks, M., Pawelzik, K., and Markram, H.: Neural networks with dynamic synapses. Neural Computation, 10:821-835, (1998).
17. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proceedings of National Academy of Sciences of USA, 79: 25542558, (1982).
18. Amit, D.: Modeling Brain Function: The world of Attractor Neural Networks. Cambridge University Press, Cambridge, (1989).
19. Poucet, B., Save, E.: Attractors in Memory, Science Perspectives, 308: 799-800, (2005)
20. Cohen, M. A., Grossberg, S.: Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Transactions on Systems, Man and Cybernetics, 13: 815-821, (1983).
21. Treves, A., Rolls, E. T.: Computational Constraints Suggest the Need for Two distinct Input Systems to the Hippocampal CA3 Network. Hippocampus, vol.2, no.2, 189-200, (1992).
22. Latham, E. P., Deneve, S., Pouget, A.: Optimal computation with attractor networks. Journal of Physiology, 97:683-694, (2003).
23. Romani, S., Tsodyks, M.: Continuous Attractors with Morphed/Correlated Maps. Plos Computational Biology, vol.6, issue.8, 1-19, (2010).
24. Zhang, K.: Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. Journal of Neuroscience, 16:21122126, (1996).
25. McNaughton, B. L., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., Kudrimoti, H., Qin, Y., Skaggs, W. E., Suster, M., Weaver, K. L.,: Deciphering the hippocampal polyglot: the hippocampus as a path integration system. The Journal of Experimental Biology, 199:173185, (1996).
26. Kali, S., Dayan, P.: The involvement of recurrent connections in area CA3 in establishing the properties of place fields: a model. The Journal of Neuroscience, 20:74637477, (2000).