Models of Associative Memory

Wenting Jin, Lucas Arnström & Robin Eklind

2016-06-03

Contents

1 Introduction

Associative memory may be defined as *"the ability to correlate different memories to the same fact or event"* [14]. Two broad categories of associative memory distinguish between memory recalled from partial information or cues (auto-associative memory), and memory recalled from related categories or concepts (hetero-associative memory). To give an example, auto-associative memory is used when asked to fill in the missing parts (e.g. *"Which country in Europe starting with an "F" is known as "the land of a thousand lakes"?"*), while hetero-associative memory is used when asked what thoughts comes to mind when presented with a given concept (e.g. *"If I say elephant, you may say pink, big or animal."*).

1.1 Aim and Objectives

The aim of the project is to investigate the current capabilities of artificial models for associative memory.

The achieve this aim, the following objectives have been identified.

1. Outline key models for associative memory from different fields of research (e.g. Hopfield, memristors, ...).
2. Evaluate the current capabilities of the models for associative memory both in terms of their capacity to store correlated information and in terms of the applications in which they've been used.

1.2 Hebbian Learning

What underlying principle or set of principles govern learning and the formation of memories? Inspired by recent discoveries in neuroscience and synaptic research, Donald Hebb proposed an iconic postulate in 1949, *"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."* which is often summarized as *"neurons which fire together, wire together"* [7]. Ever since its formulation, Hebb's rule has been a cornerstone in understanding the inner workings of learning, memory formation and associative memory. Later research has extended Hebb's rule to include a notion of decay, where synaptic connections are weakened (and subsequently associations broken). Without this addition, all synapses would eventually converge to their maximum efficiency through synaptic potentiation, thus making them in-differentiable and thereby removing the selective property which is key for associative memory recall [18].

2 Models of Associative Memory

Three models for associative memory are explored in this section, two traditional machine learning models and one emerging technology. While the traditional models (Hopfield networks and Boltzmann machines) may be considered abstract models defined independent of underlying hardware, they are in practise often employed on commodity hardware with a von-Neumann architecture. The emerging technology (memristors) however, enables a radically different computer architectures on top of which new models for associative memory (and other machine learning tasks) may be developed, and for which existing models may be ported to.

2.1 Hopfield Network

Hopfield's model is an associative(content-addressable) memory model using binary information and its network is featured with threshold and feedback. A binary input vector to a Hopfield Network will be recursively processed(fed back) at the granularity level of bit, until a threshold criterium is met, the produced output vector possesses the feature that is with the shortest Hamming distance to its origin input vector (see figure 1). There properties give rise to associative memory recall, where an input pattern (i.e. memory cue) is presented to the Hopfield network, which produces an output vector that is closest in Hamming distance to a previously stored memory, thus filling in the gaps to recall a memory through association.

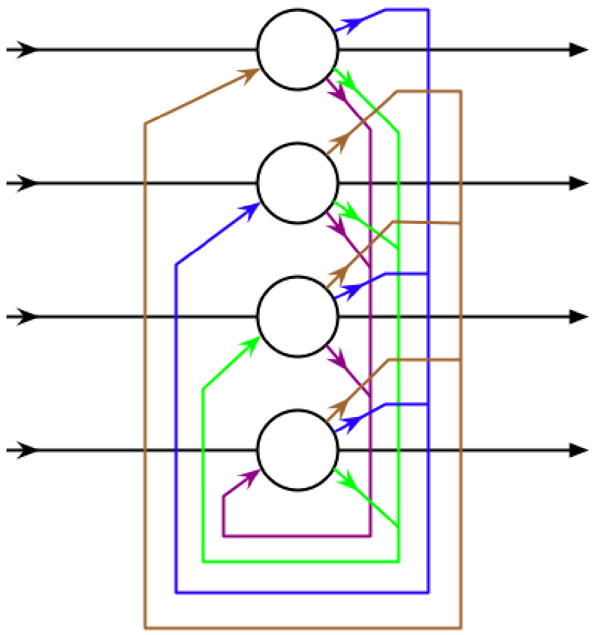

Figure 1: A Hopfield network consisting of 4 nodes, with an input and output vector of length 4. Additionally, each node takes as input the state of each other node in the network.[1]

In paper from J.J.Hopfield(1982) [9], the drawbacks of Perceptron were addressed through its intractable back-coupling, lack of abstraction properties and requirement of synchronism. Information storage was improved with help of Non-linearity, and emergent computational properties were obtained from simple properties of many cells rather than complex circuitry(which is a result of linear associativity). The input-output relationship

[1] Original image (CC BY-SA): `https://upload.wikimedia.org/wikipedia/commons/9/95/Hopfield-net.png`

of non-linear computation, binary threshold units and concept of the energy function were introduced.

Collective behaviours of the model was studied and resulted with the following findings: a few stable states was resulted from most of the initial state space, properties necessary for a physical content-addressable memory were not dependent on the symmetry of the connectivity matrix T_{ij}. Statement supported by findings from experiments is that "about 0.15N(memory storage bits, number of neurons) states can be simultaneously remembered before error in recall is severe", which refers to the memory capacity of the associative memory. Case with arbitrary starting state was studied and results of memories near to the starting state was highly produced. The nominally assigned memories which were called "attractors" dominates the phase flow whereas the flow is not entirely deterministic, which leads to a convergence to local optimum.

Case of consistent internal correlations in the memory was as well addressed, and Hebb synapses was used and slightly modified to generate non-symmetric term δT_{ij}, which limitation of sequence of four states was addressed.

2.2 Boltzmann Machine

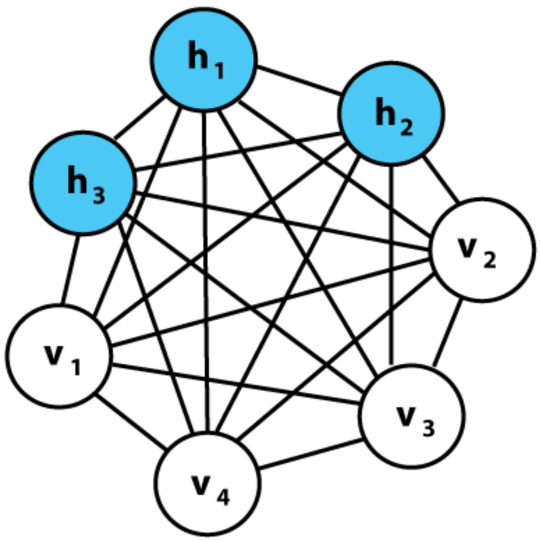

Figure 2: Illustration of a Boltzmann Machine. The blue units represents three hidden units, while the four white units represents four visible units.[2]

The "Boltzmann Machine" (BM) is a form of "parallel constraint satisfaction network" [1]. It is capable of learning the underlying constraints of a domain by only being shown examples of it. The BM is composed of units (also known as nodes) forming a complete graph where the connection between two units are symmetric; meaning that the weight on the connection is the same in either direction. No unit has a connection to itself. The units are binary, in the meaning that they can only assume one of two states, on or off. The state of a unit is determined by a probabilistic function based on the states of the

[2]Original image (Public Domain): `https://en.wikipedia.org/wiki/File:Boltzmannexamplev1.png`

units neighbours. A strong connection (high weight value) between two units indicates that if either of these two units are active, the other one should probably be active as well. While a weak connection (low weight value) indicates that these should probably not be active at the same time. This is analogous to Hebbian learning.

The BM is notably similar to the Hopfield network in that it also defines a global energy state of the system, utilizing the same equation that determines the global energy value. Each global state can be identified by the energy of the system in that state. By forcing the values of the visible units to represent a training set the system attempts to find an energy configuration that is compatible with the given input. The resulting energy state can then be interpreted as to how well the given data fulfils the constraints of the domain. Thus by minimizing the energy the system learns an interpretation of the problem that increasingly satisfies the constraints of the domain.

The simplest way to minimize the energy into a local minimum of the system is to change each units state into a value that results in a lower energy state. The data needed to determine this change is locally accessible to each unit, and is dependent on the current state of the units neighbours. If the sum of all values for a given units neighbour exceeds the threshold of that unit, the resulting state of the unit should be on. Otherwise it should be off. This is the usual algorithm for binary units.

Because of this deterministic algorithm it suffers from the usual weaknesses of gradient descent algorithms, namely, it gets stuck in local minima if its initial state is close to one. In order to alleviate the algorithm of this problem, noise is introduced in the training. This allows the network to "jump" out of these minima into configurations of higher energy. The algorithm used for noise introduction is a variation of the "Metropolis algorithm" [11] that was used to study thermodynamic systems. This modified version introduces a concept of temperature to the machine, which then tries to reach "thermal equilibrium" during training. Meaning that the machine is allowed to run repeatedly until the global energy of the system converges to a fixed state over a temperature that is initially high and then slowly decreased over the runtime of the system. The probability of finding the system in a global state after it has reached thermal equilibrium follows a Boltzmann distribution. If the temperature feed into the machine is equal to zero the stochastic nature of it is removed, the machine becomes deterministic and can be seen as a regular Hopfield network.

Training is conducted in two phases, in the first phase the visible units of the machine are set to the values of the training set. In the next phase the machine is allowed to run freely, independent of the training set. The machine is iteratively switched between these two phases for the duration of the training until it reaches thermal equilibrium. The goal of the training is for the machine to be able to generate the input vector with a high probability of success.

2.3 Memory Resistor

In 2008, a group of researchers from HP labs published a paper entitled *"The missing memristor found"* [21]. What, then, is a memristor, and how did we know it was missing? A memristor (short for memory resistor) is a resistor with memory whose resistance depends on the past flows of current passed through the circuit. The resistance of a memristor is increased by current travelling through it in one direction, and decreased by current travelling through it in the other direction. A memristor is a passive circuit which remembers its resistance even when inactive and without power for long periods of time. These properties make memristors interesting candidates for non-volatile storage and memory units, as they retain their state when unpowered, and specifically as they enable storage of continuous ranges of values (i.e. low *through* high resistance) in contrast to discrete binary values (i.e. 0 *or* 1) [12].

Back in 1971, Leon Chua, often referred to as the father of non-linear circuit theory, laid the mathematical foundation detailing the relations between the four fundamental circuit elements. Interestingly, at that time only three fundamental circuit elements had physical counterparts, namely resistors, capacitors and inductors. The fourth fundamental circuit element, the memristor, was only conceptualized in theory by Chua for the sake of symmetry (see figure 3). As outlined in Chua's seminal paper *"Memristor-the missing circuit element"* [3], the current-I voltage-V curve of a memristor has a unique shape, an IV-fingerprint if you will, in the form of a pinched hysteresis loop (see figure 4). A hysteresis loop indicates that a system has an internal state (i.e. memory) which affects the output of the system and which depends on past inputs to the system [6]. As famously state by Chua, *"If it's pinched it's a memristor"* indicates that the hysteresis loop of a memristor passes through the origin.

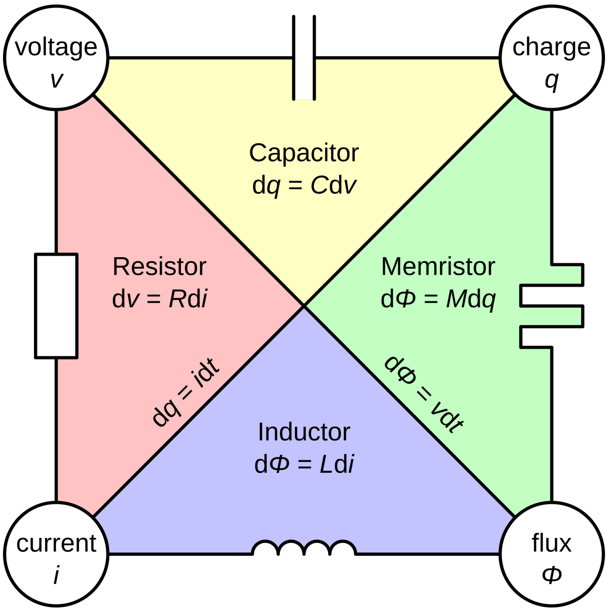

Figure 3: Fundamental circuit elements.[3]

Ever since 2008, there has been an exponential increase in research related to memristors, where the number of search results for "memristor" on Google Scholar has doubled every 18-24 months [12]. Why are so many researchers attracted to this new field of research? To answer this question, lets first evaluate the suitability of using current computer architectures in highly adaptive systems, such as brain-like learning systems with neural and synaptic placticity (i.e. adaptivity).

[3]Original image (CC BY-SA): `https://en.wikipedia.org/wiki/File:Two-terminal_non-linear_circuit_elements.svg`

[4]Original image (© Brian Hayes): `https://www.americanscientist.org/libraries/documents/201128120228377-2011-03CompScienceHayes.pdf`

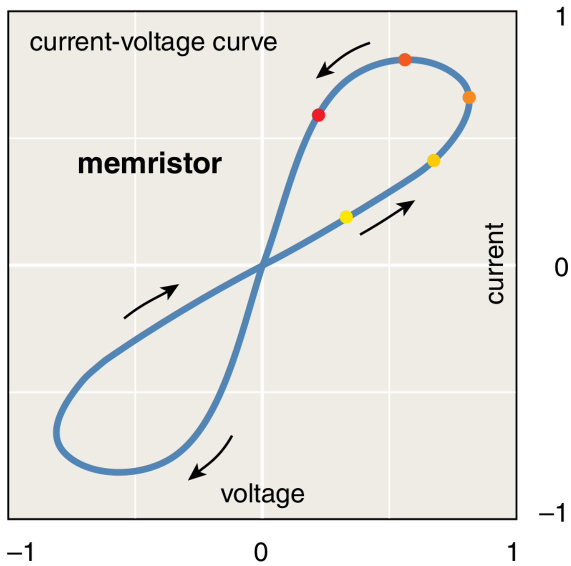

Figure 4: IV-curve of a memristor circuit, arrow indicates time.[4]

Machine learning research has managed to achieve some truly remarkable milestones in recent years (e.g. AlphaGo beating the human world champion in Go [19]), both reaching and surpassing human potential on a number of tasks for which the brain is tailored towards, such as face recognition [17], classification, and abnormality detection. Given this resent development, it may be tempting to imagine that it is only a matter of time until these machines achieve general problem solving skills through highly adaptive learning, which is fundamental for general artificial intelligence.

However, the very nature of adaptivity introduces a significant challenge for today's computer architectures as it is inherently dependent on mutable states to reflect changes in the environment. To model these changes in state, data has to be shuffled back and forth between the processor and memory. The separation of processing and memory access is the underlying cause of two significant problems with current computer architectures. Firstly, it restricts the potential for parallel computation [2] and introduces a set of complex workarounds (e.g. mutual exclusion, cache line invalidation). Secondly, and perhaps more importantly, a substantial amount of energy is required just to shuffle data back and forth between the processor and memory [13].

There exist a huge discrepancy between the energy requirements of adaptive learning systems implemented in nature by biological brains and those implemented in silicon by machine learning algorithms running on von-Neumann computers. The difference in energy efficiency for adaptive learning tasks is estimated to be around 9 orders of magnitude. To put this into perspective, the brain would be able to travel around the entire earth on the same amount of energy that current computers would require to travel one and a half inches [12]. This is one compelling reason why researchers are interested in understanding the inner workings of the brain, so that fundamental principles for energy efficient adaptive learning may be derived and modelled.

2.3.1 Artificial Synapses

Several similarities have been identified between the properties of memristors and synapses, which make them interesting candidates for associative memory models; as further described in section 3.4.

3 Current Capabilities

The current capabilities of the three different models are described both in terms of their capacity to recall associative memories, and in terms of concrete applications for which the models have been used.

3.1 Hopfield Network

3.1.1 Implementation Example: Optical Information Processing

In paper from D. Psaltis and N. Farhat(1985) [15], threshold and feedback properties from Hopfield model was used in implementation of an optical system that processes optical information.

Enhanced error-correcting capability is one important outcome which is benefited from Non-linearity of Hopfield model. With M words of each binary vector v_i with length of N bits, matrix T_{ij} represents a storage of information. Matrix T_{ij} gets multiplied by a stored binary vector v_i' results in a `pseudoeigensystem` if N is sufficiently larger than M. This indicated that the output vector v_i'' equals the input.

Supposed non-linear iterative procedural experiments of certain number of known bits N_1 and the rest bits set to zero inside total of N bits long vector were addressed to discover under what conditions the number of correct bits N_2 in output will be higher than N_1. A SNR(signal-to-noise ratio) equation which consists ratio of the expected value to the standard deviation on the same output vector, was used. As the Hopfield model has studied on the convergence property with respect to asynchronous operations, insensitivity to imperfections(non-uniformities, exact form of the threshold operation and errors in T_{ij} matrix) and correct convergence obtained with threshold T_{ij}. These all become most desired properties in optical implementation. Detailed optical implementation with 2D inputs was presented which was based on spatial-frequency multiplexing. Methods using Fourier transform, transmitting amplitude(weighted sum) and integral of the product of the input images were introduced in such an implementation. The robustness of a such system with non-linear feedback becomes the most important feature. As a conclusion, the implementation with the capabilities and limitations of optical techniques matches excellently with the Hopfield model that requires global, linear operations and local, point non-linearities in a fully interconnected optical system.

3.1.2 Memory Capacity with Modification: replacing sigmoid neuron with a non-monotonic neuron

In paper from S. Yoshizawa, M. Morita and S. Amari(1992) [22] it started with introduction of a new method by replacing sigmoid neuron with a non-monotonic neuron and discussed theoretically on potential of absolute capacity(the maximum number of randomly generated patterns which are memorized as the equilibria of the network with the correlation-type connection weights) to be of order n (nearly equal to $0.4n$).

Previous memory capacities were briefly introduced that Hopfield(1982) model's associative memory capacity is $0.15n$, the proven result of absolute capacity is asymptotically $\frac{n}{2\log n}$ (from McEliece, Posner, Rodemich, & Venkatesh, 1987; Weisbuch, 1985), and relative capacity(recalling process) is about 0.14n(with admission of small percent of errors) with replica method, but about $0.16n$ with a simply approximation method from respectively two different research work.

Various research suggestions were mentioned though all failed to deal completely with flaws of the conventional model, that both absolute and relative capacities are too small as well as the existence of large number of spurious memories. With a non-monotonic neuron the result turned to be $0.4n$ for the absolute capacity which is even greater than relative capacity while the spurious memory disappeared.

With conventional neuron, memorized pattern is unstable, whereas a basin of attraction around memorized pattern was shown with non-monotonic neuron by replacing sigmoid function with a non-monotonic output function *figure 5* to neuron elements from the recalling process of an autocorrelation associative memory. This was named as Morita Model in the paper though it is essentially an extension from Hopfield Model. By then the existence of equilibrium solutions and local stability, the authors did not further investigate on problems as the following: the size of the basin of attraction, the full sketch of spurious memories and the behaviour for clustered memorized patterns, which leaves more research work for future researchers to understand better associative memory with non-monotonic neurons.

3.2 Experiment on Implicit Memory for Novel Associations between pictures: Effects of Stimulus Unitization and Aging

From various previous research, concepts such as associative priming, unitization, difference between conceptual and perceptual associative priming, verbal versus pictorial material/stimuli and roll of spatial proximity were briefly summarized [10].

Experiments with pictorial stimuli(paired pictures) were done in 3 consecutive stages, where the result from first stage showed no evidence on requirement of spatial contiguity, though associative priming was enhanced compared to with spatially separated stimuli, which proved "implicit memory for novel associations still can occur in the absence of an emergent conceptual representation". The second experiment was an extension from the first experiment, with focus on the effects of aging and spatial contiguity of the same topic of stimuli on novel association priming between pictures, where stunning result

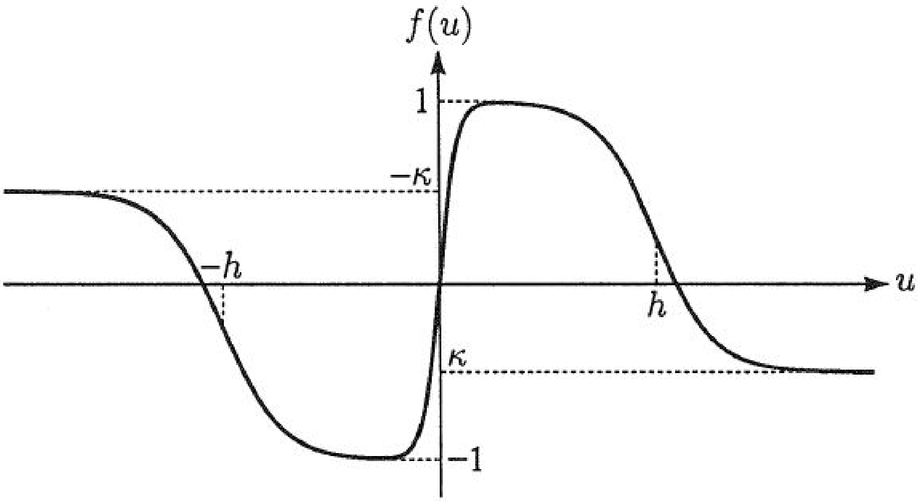

Figure 5: The non-monotonic output function, S. Yoshizawa, M. Morita and S.Amari(1992)

was shown that “associative priming is age invariant”(exposure of pictures was longer with older group to yield a matched performance in the baseline). The last experiment was based on both first and second experiments that “associative priming with pictorial stimuli is modulated by spatial contiguity but not by aging”, and the study proved further evidence for the notion that novel association priming for picture pairs is mediated by the PRS(Perceptual Representation System).

3.3 Boltzmann Machine

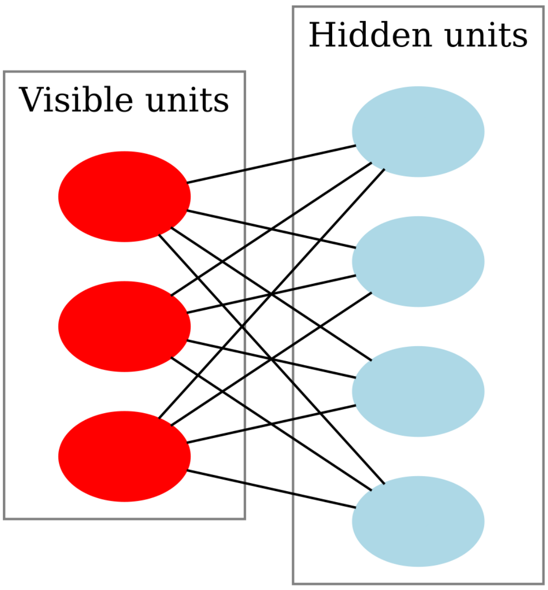

Figure 6: Illustration of a Restricted Boltzmann Machine.[5]

Unfortunately, the Boltzmann Machine is very slow to train and is thus only practical for simpler problem domains. If the BM is scaled beyond any trivial domain it becomes too slow and almost stops learning. This is due to the fact that all units of the BM are fully connected to each other which does not scale well.

In order to use the BM for bigger tasks a restriction has to be made. Namely, connections between units in the same layer can not be allowed. This is called the Restricted Boltzmann Machine (or "Harmonium" as the original author referred to it) [20].

Versions of the Restricted Boltzmann Machine (RBM) have been successfully used in many applications such as deep learning [8] and speech recognition [4].

The general idea when using RBM's in deep learning is to "stack" several RBM's on top of each other. The activities of the units in the hidden layer of one RBM can be used as input vector for the next RBM. This way the overall system does not have the problem with scalability that the ordinary Boltzmann Machine suffered as well as the added bonus that the generative model is improved each time a new layer is added on top of the existing ones.

A variant of the Deep Boltzmann Machine [16] (which is used for deep learning) called the Shape Boltzmann Machine (SBM) has been shown to be able to "restore" parts of an image when shown only a part of it [5]. It does not "restore" the image to its original form, but fill in the blanks with its interpretation of the data it has been given. This can bee seen as a form of auto-associative memory.

3.4 Memory Resistor

Several independent research groups have been exploring how to make use of memristors to build artificial synapses [13, 14]. The efficiency of a synapse may essentially be encoded as the resistance of a memristor, and synaptic networks may thus be built using several memristors, where neural activity will strive for the path of least resistance.

An experiment was conducted in 2010, which used memristors as artificial synapses to empirically validate the formation of associative memories between three neurons and two synapses (see figure 7). In the experiment, two input neurons and one output neuron was used. The input neurons were connected to the output neuron, using one artificial synapse each. When the input neurons were triggered at the same time, their activity became associated to each other, as the resistance of their memristors decreased. This made it possible to activate one neuron, simply by activating the other, which may be compared to the classical experiment by Pavlov where dogs were conditioned to salivate when presented with triggering stimuli, e.g. the sound of a bell.

[5] Original image (CC BY-SA): https://en.wikipedia.org/wiki/File:Restricted_Boltzmann_machine.svg

[6] Original image (© Yuriy V. Pershin and Massimiliano Di Ventra): https://arxiv.org/pdf/0905.2935.pdf

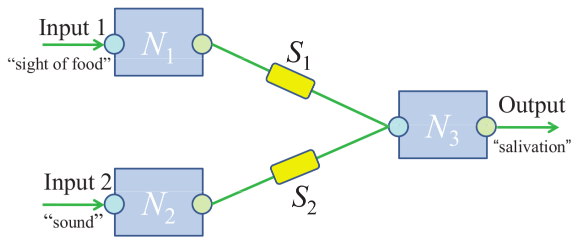

Figure 7: Associative memory formation (in the form of traditional conditioning) achieved using artificial synapses.[6]

4 Conclusion

All three models make use of Hebbian learning to achieve formation of associative memories. Models such as the Hopfield network and the Boltzmann Machine provide key insight into verifiable theories behind the formation of associative memories. While their theories today is nowhere near what any living brain is capable of their usage in experimentation and further studies have helped our understanding of how human like memory can be implemented in software.

Since both the Hopfield network and the Boltzmann Machine are mainly theoretical models implemented on top of commodity hardware, their performance and energy usage will never get close to that of something completely designed in hardware bottom up, such as the Memory Resistor. While the memristor is still in early development it shows promising capabilities regarding these issues.

References

[1] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for boltzmann machines," *Cognitive science*, vol. 9, no. 1, pp. 147–169, 1985.

[2] K. Boahen, "A computer that works like the brain," TED talk, june 2007, [Online] Available: https://www.ted.com/talks/kwabena_boahen_on_a_computer_that_works_like_the_brain. [Accessed: 9 may, 2016].

[3] L. O. Chua, "Memristor-the missing circuit element," *Circuit Theory, IEEE Transactions on*, vol. 18, no. 5, pp. 507–519, 1971.

[4] G. Dahl, A.-r. Mohamed, G. E. Hinton *et al.*, "Phone recognition with the mean-covariance restricted boltzmann machine," in *Advances in neural information processing systems*, 2010, pp. 469–477.

[5] S. A. Eslami, N. Heess, C. K. Williams, and J. Winn, "The shape boltzmann machine: a strong model of object shape," *International Journal of Computer Vision*, vol. 107, no. 2, pp. 155–176, 2014.

[6] B. Hayes, "The memristor," *American Scientist*, vol. 99, no. 2, pp. 106–110, 2011.

[7] D. O. Hebb, *The organization of behavior: A neuropsychological approach.* John Wiley & Sons, 1949.

[8] G. E. Hinton and R. R. Salakhutdinov, "A better way to pretrain deep boltzmann machines," in *Advances in Neural Information Processing Systems*, 2012, pp. 2447–2455.

[9] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.

[10] I. P. Kan, M. M. Keane, E. Martin, E. J. Parks-Stamm, L. Lewis, and M. Verfaellie, "Implicit memory for novel associations between pictures: effects of stimulus unitization and aging," *Memory & cognition*, vol. 39, no. 5, pp. 778–790, 2011.

[11] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[12] T. Molter, "Memristor technology - a new and exciting frontier," in *Workshop on Memristor Technology, Design, Automation And Computing (MemTDAC) Affiliated with the HiPEAC 2016 conference, (Jan 2016)*, 2016.

[13] M. A. Nugent and T. W. Molter, "Ahah computing–from metastable switches to attractors to machine learning," *PloS one*, vol. 9, no. 2, p. e85175, 2014.

[14] Y. V. Pershin and M. Di Ventra, "Experimental demonstration of associative memory with memristive neural networks," *Neural Networks*, vol. 23, no. 7, pp. 881–886, 2010.

[15] D. Psaltis and N. Farhat, "Optical information processing based on an associative-memory model of neural nets with thresholding and feedback," *Optics Letters*, vol. 10, no. 2, pp. 98–100, 1985.

[16] R. Salakhutdinov and G. E. Hinton, "Deep boltzmann machines," in *International conference on artificial intelligence and statistics*, 2009, pp. 448–455.

[17] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815–823.

[18] T. Sejnowski, "Statistical constraints on synaptic plasticity," *Journal of theoretical biology*, vol. 69, no. 2, pp. 385–389, 1977.

[19] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[20] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," DTIC Document, Tech. Rep., 1986.

[21] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *nature*, vol. 453, no. 7191, pp. 80–83, 2008.

[22] S. Yoshizawa, M. Morita, and S.-I. Amari, "Capacity of associative memory using a nonmonotonic neuron model," *Neural Networks*, vol. 6, no. 2, pp. 167–176, 1993.