A Summary of “The Capacity of the Hopfield Associative Memory”

The University of Waterloo - ECE657- A2Q2
by Pawel Jaworski, Ted Themistokleous

The paper “The Capacity of the Hopfield Associative Memory” provides a rigours analysis and then proof of the overall theoretical capacity of a Hopfield artificial neural network. The work starts with a formalized recantation of Hopfield’s original work, and initially discuses desired operation, and formulation as well as assumptions. The later sections the paper delve into adjusting and generalizing original work to provide a more rigorous proof of the following; Convergence towards stored vectors (called fundamental memories), and gives a probabilistic upper and lower bound on the amount of fundamental memories are able to be stored within such a network. The heuristic of recall of some, or all the memories is also discussed and used in the formulation of these bounds. This approach invokes various probabilistic theorems, and also culminates in the final sections of the paper while coming to the conclusion of an upper bound for possible memory capacity with respect to vector length as the number of fundamental memories.

The initial introduced formulation discusses the idea of a vector of bipolar values (-1/+1) of size n ‘entries’, and a fixed set of m vectors that are used to build the initial encoding for storing the m-th entry. The main assumption in this section is that m, the number of different fundamental memories, is much smaller than the entry vector size n, allowing an external new unknown pattern, called the ‘probe’ to recall one of these memories. Formulation of the overall synaptic weight matrix T is then shown to be the sum of the following; Take the outer product each fundamental memories and then subtract the identity matrix, leaving each memory with a zero diagonal in the resulting matrix. T, is then used with the input probe, and all values are fed through a signum function of the result giving a positive or negative flip of the resulting state for a randomly chosen bit. The idea if each “bit” update being either synchronous or asynchronous are discussed but assumed to have a negligible difference in performance on each bit update as the author describes as the network “homes in” to a fundamental memory stored in the system. This is used to begin asking why the system converges to a fundamental memory, but not specifically the correct one. A specific example is shown that yields differing results of output bits based on this notion as a way to use convergence as a means to determine heuristically the upper limit of capacity.

Stability of this algorithm is then discussed as the use of a probabilistic approach beings to take shape. Each fundamental memory modelled as a sphere in a larger Hamming space where each of the m spheres, have a radius p * n where, p is the number of 0 to 0.5 to represent up-to 50% of the “wrong” moves towards the center of a each sphere. The radius is called the radius of attraction and is used throughout the later proof to determine the capacity of the network based on different cases of attraction towards the fundamental memory and the edge of the boundary. The paper then begins discussing the actual capacity heuristics by modelling each fundamental memory breaking things into two main cases; every fundamental memory may be fixed at a sphere center and having a radius of attraction while the weaker concept where “almost every memory is good” thus leading to double the capacity from the later proofs. Capacity discussed also implies that the value of p for the radius of attraction stay fixed. The positive and negative components in each item in the set of fundamental memories is then shown to follow a standard normal variable for positive components and poison for negative components and this is used as a base point for which the formalized proofs rely on.

The next section in the paper begins introducing four lemmas, relying on the fact that all random variables used to model each fundamental memory are unbiased. Lemma A uses the large deviation theorem in that each probe used in the proof can be withing some order (denoted O(N ^ 3/4)) away from the mean of each fundamental memory center before some sort of convergence. This differs

as the initial formulation used the notion of a probe needing to be within p*n of a memory to converge within the hamming space. Lemma B and C are build of Lemma A and used to later prove the larger theorem. Lemma B is uses lemma A to get a rate-of-growth based expression for the set of n (the size of the vector) random variables,where as lemma B' invokes a "strong form of the large-deviation central limit theorem" with lemma C using Bonferroni's inequality to base the next more rigorous section.

Once the four lemmas are introduced a more generalized and rigorous approach is taken with the formulation with the weight matrix, such that each fundamental memory outer product is not subtracted by the Identity matrix, but a varying delta*g value that can vary from [0, 1] for the storage. The rigorous proof uses the lemmas and the generalized formulation and explores the cases of convergence with varying correct, and incorrect changes in attraction to the fixed centers with a growth bound of the value m with respect to bit width of each memory vector used. The sum of each of the components is used (denoted Sj) and is used throughout the proof to check the probability of whether a bit flip will occur given correct and incorrect changes in probe vector with respect to a fundamental memory stored in the weight matrix. The later proofs utilize the previous sections lemma A, B, B',C on the sum of the entries to trigger a bit flip. A bi-part graph is used to simplify and explain the fixed point fundamental changed between each move and shows that no cyclical paths must be taken between moves in the sums trajectory for each probe to converge to a fundamental memory in the specified radius of attraction. As a result the notion of $((1 - 2p)^2)/4 * (n / \log n)$ for indirectly attracted to a fundamental memory, and $((1-2p)^2)/ 2 * (n / \log n)$ for a direct convergence to the fixed center asymptotic bound for capacity.

The final sections informally extend the rigorous proof sections, by asking the question " what happens if wrong changes occur during the convergence of a probe to a fundamental memory". It is then shown for both the synchronous and asynchronous update cases for a small error (1 - e) and e > 0, how convergence changes with respect to fixed points. This result is used for the justification to remove the $(1-2p)^2$ factor which then yields the following result; Should probe be some p*n away from a memory, with p bounded [0, 0.5) of the Hopfield network can be designed to support an asymptotic capacity of $m = n / (2 \log n)$ should the need to remember only all but a few of the encoded memories is desired and, $m = n / (4 \log n)$ if all memories need to be remembered. It is also shown that as the radius increases (p -> 1/2), n the number of entries in each memory vector, needs to be larger and as this implies a probe would make more errors to converge to the a resulting fundamental memory. The $m = n / (4 \log n)$ result would provide a complete upper bound in storage should all memories be required to be remembered in this case and allow for wrong moves in the network as the system reached a stabilized point, found in the paper to be around two wrong moves regardless of the synchronous or asynchronous update.