Spreading Activation and Sparseness in a Bidirectional Associative Memory

C. Tremblay, M. Dorville, K. Myers Stewart, and S. Chartier

Abstract—The Bidirectional Associative Memory (BAM) is a type of artificial neural network that was shown to have great performances in learning and recalling various types of associations. However, this model has always been investigated under optimal conditions in which all the patterns have the same desirability and the network is fully connected. In this paper, the influence of spreading-activation and sparseness in a BAM network is studied. Results show that even under such variability the performances of the BAM are unaffected. This study gives us a better understanding of how attractors can be developed and could lead to more robust computational intelligence systems.

I. INTRODUCTION

B eing able to recognize and recall patterns of various natures and different contexts is something that human beings routinely accomplish with little effort. However, these tasks are difficult to reproduce through the use of artificial intelligence systems. A successful approach in the past has consisted of distributing information over parallel networks of processing units, as is done in biological neural networks. Brain inspired recurrent associative memories offer the ability to develop attractors for each pattern through feedback connections such as the Hopfield model [1]. Kosko [2] later generalized the Hopfield model to a heteroassociation, thereby creating a new class of neural network models, the bidirectional associative memory (BAM). Numerous BAM models have been developed since Kosko’s (e.g. [3], [4], [5], [6]). One model proposed by Chartier & Boukadoum [7], [8] was the introduction of a unique matrix for each layer. This Bidirectional Heteroassociative Memory (BHM) is able to learn correlated patterns for bipolar patterns as well as for real-valued patterns. Although the BHM network exhibits properties that make it cognitively interesting, its learning has always been performed in an optimal setting; where each pattern has the same importance and the units are fully connected. In other words, during the learning phase, no variability was taken into account.

Much research in the field of cognition has been done on variability added to neural connections in function of concept's relatedness. An example of this phenomenon is that people are faster at recalling an image of an apple or the word “apple” if they have been primed with associations of images or names of fruits. A commonly cited work by Collins & Loftus [9] proposed the spreading activation theory of semantic processing, which is an extension of Quillian's theory of semantic memory search and semantic preparation [10], [11]. The theory suggests that the activation of a semantic concept in the brain, by priming for example, spreads across neighbouring concepts, resulting in the lowering of the activation threshold for neighbouring concepts. The spreading activation would then dissipate as a function of semantic distance, which is supported by empirical research [12]. This theory was later applied to memory where it was said that the probability of learning a pattern is a function of the pattern's relatedness with neighbouring concepts [13]. Taking into account spreading activation in models has proven useful for information retrieval techniques where performances are higher when the activation of a concept spreads to related concepts [14], [15]. In the field of connectionist models, research has showed that the use of spreading activation does not affect the convergence of feed-forward networks [16] and word processing neural networks [17]. In other words, various settings still converge to a stable solution. Although the application of spreading activation has shown promising results in artificial neural network models [16], [17], the application of such variability has not yet been tested on the BAM.

Another issue related to the variability of the activation is the case of sparseness. Sparseness can be seen as a special case of variability when the activation is null. A body of evidence from neuroscience has shown that the brain is sparsely connected [18], [19], [20], [21], [22], [23], [24], unlike most artificial neural networks where the information is fully distributed. In the brain, neural connections require energy and space, making a fully connected system biologically costly and unnecessary [25]. Successful uses of sparse connectivity were done in connectionist models and have shown that sparse connectivity still leads to stable solutions [26], [27], [28], [29]. In addition, sparse connectivity shows increased capacity to generalize in feed forward neural networks [30], and increased storage capacity in some associative memories [25], [31], [32]. Again, the BAM models do not take into account a certain level of sparseness.

In short, for both the spreading activation as a function of semantic distance and sparseness connections, the process is done through a modification of the learning parameter. In fact, instead of having a unique parameter for the whole network as in the BAM, it is proposed that the network uses...
a matrix of parameters instead. In other words, one learning parameter is used per unit. If the values of many learning parameters are null, then the network acts like a sparse network. If the values are not null but different from each other, then the network takes into consideration various semantic distances. Finally, if the parameters are not null but equal, then the network behaves in standard fashion.

This paper therefore seeks to evaluate the proprieties of the BHM when spreading activation is added to the patterns to process as well as when the network is sparsely connected. First, simulations on the BHM network introduced in Chartier & Boukadoum [7], [8] will test the influence of spreading activations. Second, simulations on the same network with different levels of sparseness will be performed to evaluate the robustness of the BHM regarding its attractors to different levels of connection sparseness.

The remainder of the paper is divided as follows: Section II describes the network's architecture. Section III shows simulation results regarding the network's performance in learning and recalling: 1) with variability added; and 2) with different levels of sparseness. Section IV discusses the results and provides a conclusion of our work.

II. MODEL

The network's architecture is shown in Fig.1 where \( \mathbf{X}(0) \) and \( \mathbf{Y}(0) \) represent the initial vectors-states, and \( \mathbf{W} \) and \( \mathbf{V} \) are the weight matrices and \( t \) is the current iteration number. The network is made of two Hopfield-like neural networks interconnected in a head-to-tail fashion, providing a recurrent flow of information that is processed bidirectionally.

![Fig. 1. Architecture of the BHM.](image)

A. Transmission function

The output function is defined by the following equations:

\[
\forall i, \ldots, N, y_{i[t+1]} = f(a_{i[t]})
\]  
\[
= \begin{cases} 
  1, & \text{if } a_{i[t]} > 1 \\
  -1, & \text{if } a_{i[t]} < -1 \\
  (\delta + 1)a_{i[t]} - \delta a_{i[t]}^3, & \text{else}
\end{cases}
\]  

(1b)

where \( N \) and \( M \) are the number of units in each layer \( i \) is the unit index, \( \delta \) is a general transmission parameter, \( a \) and \( b \) are the activation. These activations are obtained the usual way: \( a(t) = \mathbf{W}x(t) \) and \( b(t) = \mathbf{V}y(t) \). Figure 2 illustrates the shape of the transmission function for \( \delta = 0.2 \). In short, the equation consists of a sigmoid-type function with two hard limits added at 1 and -1 instead of being asymptotic. The transmission function has the advantage of exhibiting grey-level attractor behaviour that contrasts with other BAMs that can only develop bipolar attractors [2].

![Fig. 2. Transmission function for \( \delta = 0.2 \).](image)

B. Modification of the learning parameter.

Usually in the BHM and other BAMs, the solution is found iteratively by applying a small constant to weight updates. In order to implement sparseness and spreading activation, the learning parameter has been replaced by a matrix

\[
A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\
         a_{21} & a_{22} & \ldots & a_{2m} \\
         \vdots & \vdots & \ddots & \vdots  \\
         a_{n1} & a_{n2} & \ldots & a_{nm} 
\end{bmatrix}
\]

for one layer and

\[
B = \begin{bmatrix} \beta_{11} & \beta_{12} & \ldots & \beta_{1n} \\
         \beta_{21} & \beta_{22} & \ldots & \beta_{2n} \\
         \vdots & \vdots & \ddots & \vdots  \\
         \beta_{m1} & \beta_{m2} & \ldots & \beta_{mn} 
\end{bmatrix}
\]

for the other layer. Therefore, each connection can be modified independently from each other and each layer has its own matrix of values, different from one another.

C. Learning rule

The weight connections are modified following a Hebbian/anti-Hebbian approach [33], [34]:

\[
\mathbf{W}_{(k+1)} = \mathbf{W}_{(k)} + A(\mathbf{y}_0 - \mathbf{y}(t))(\mathbf{x}_0 + \mathbf{x}(t))^T
\]

(2)

\[
\mathbf{V}_{(k+1)} = \mathbf{V}_{(k)} + B(\mathbf{x}_0 - \mathbf{x}(t))(\mathbf{y}_0 + \mathbf{y}(t))^T
\]

(2)
where \( \mathbf{x}(0) \) and \( \mathbf{y}(0) \) are the initial inputs to be associated, A and B are the matrix of learning parameter, and \( k \) is the learning trial number. Equation (2) shows that the weight matrices will converge only when \( \mathbf{x}(t) = \mathbf{x}(0) \) or \( \mathbf{y}(t) = \mathbf{y}(0) \). Thus, each weight matrix converges when the feedback is equal to the initial inputs. In the BHM, the network convergence is guaranteed if the learning parameters \( (\alpha_{ij} \wedge \beta_{ij}) \) are smaller than the threshold found with (3) [7]:

\[
(\alpha_{ij} \wedge \beta_{ij}) = \frac{1}{2(1-2\delta)\text{Max}[M,N]} \delta \neq \frac{1}{2} \tag{3}
\]

where \( M \) and \( N \) are respectively the dimensionality of the input and its association.

III. Simulations
The simulations are divided in two sections. The first set of simulations studies the performance of the BHM on a one-to-one association task when various spreading activations are added to the network, while the second set studies the performance of the BHM on the same task when the network is instead sparsely connected. The performances of the BHM under variability are compared with the performances of a standard fully connected BHM.

A. Simulation 1: Learning and recall of one-to-one association Spreading activation
The goal of the simulation was to test the robustness of the BHM under various values of spreading activation during learning of a one-to-one association.

1) Methodology
The BHM used was described in section II, with an adjustment made to its learning parameter. For the first set of simulations, the values for \( \mathbf{A} \) and \( \mathbf{B} \) are picked randomly from a uniform distribution with the upper bound set according to Equation (3). The lower bound is obtained by reducing the value of the connections by a factor of 1.2, 3 and 10. The lower bound controls for the maximum distance between two values. In that sense, a factor of 1.2 allows for a concept to be at a maximum distance of 1.2 times another concept. Figure 3 presents a table of the possible lower and upper bounds given by the uniform function as a function of the variability given and the dimensionality. The task performed was an auto association of random binary vector patterns. The dimensions were varied from 100 to 500. Moreover, the memory load (number of inputs compared to their dimensionality) was also manipulated. The network was tested under low (7%), medium (15%) and high (30%) memory load. When the effect on the dimensionalities was studied, the memory load was set to 15% and when the effect on the memory load was studied, the dimensionalities were set to 300. In a Kosko-type BAM the maximum memory load is at about 15%. The stimuli were generated randomly and no pattern appeared more than once in a list of associations. The transmission function parameter \( \delta \) was set to 0.2 throughout the simulations and the number of iteration through the network before the weight matrices are updated was set to \( t = 1 \).

<table>
<thead>
<tr>
<th>Variability ( (i) )</th>
<th>Dimensions</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon_1 )</td>
<td>100</td>
<td>0.0069</td>
<td>0.0083</td>
</tr>
<tr>
<td>( \epsilon_2 )</td>
<td>300</td>
<td>0.0023</td>
<td>0.0028</td>
</tr>
<tr>
<td>( \epsilon_3 )</td>
<td>500</td>
<td>0.0014</td>
<td>0.0017</td>
</tr>
<tr>
<td>( \epsilon_4 )</td>
<td>100</td>
<td>0.0028</td>
<td>0.0083</td>
</tr>
<tr>
<td>( \epsilon_5 )</td>
<td>300</td>
<td>0.00093</td>
<td>0.0028</td>
</tr>
<tr>
<td>( \epsilon_6 )</td>
<td>500</td>
<td>0.00056</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

Fig. 3. Lower and upper bounds for the uniform distribution for \( i \) set to 1, 2, 3 and 10.

Learning was carried out according to the following procedure:
1) Random selection of a pair of patterns \( (\mathbf{x}(0) \text{ and } \mathbf{y}(0)). \)
2) Computation of \( \mathbf{x}(t) \) and \( \mathbf{y}(t) \) according to the transmission function (1).
3) Computation of the weight matrices update according to (2).
4) Repetition of steps 1) to 3) until all of the pairs have been presented.
5) Repetition of steps 1) to 4) until the mean square error is lower than \( 10^{-5} \).

Following the learning phase, the network was tested on a recall task under various types of noise. The network was exposed to 1) additive normally distributed vector noise, by adding to the input pattern a vector composed of random valued selected from a normal distribution with a mean of 0 and a standard deviation that varied from 0.5 to 2.5. 2) by flipping a proportion of pixels of the input pattern. In this last case, the proportion flipped varied from 10% to 50%; 50% being the maximum possible proportion for this type of noise. Each recall was performed 200 times. The mean performances are reported in the following section.

2) Results
Figure 4 presents the performances on the recall task with spreading activation. Graphs 4a to 4c show that the variability in the learning parameters does not affect the recall performances with additive random noise under a memory load of 15%. The performances remained perfect independently of the dimensionality. Similar results about pixels flipped are shown in the graphs 4d to 4f where the performances of the network are not affected by the level of strength discrepancy between each learning parameters. In other words, the performances of the network, whether or not it uses a fixed learning parameter, remained the same. In short, various levels of spreading activation will not alter the performance of the network. The small observed variations of performances are the consequence of the lower values of the learning parameters.

As for the performances under different levels of memory load in a network of 300 dimensions, results show that as the variability in \( \mathbf{A} \) and \( \mathbf{B} \) increases, the performances of the network remain very similar regardless of the level of
strength discrepancy. As is to be expected, the performances of the network decrease as the memory load is increased. The performances of the network with spreading activation remain similar to the performances of the standard BHM. Again, the small observed variations of performances are the consequence of the lower values of the learning parameter.

### B. Simulation 2: Learning and recall of one-to-one association with sparseness levels

The purpose of the simulations was to evaluate the impact of sparseness on the memory capacity and performance on an auto associative task.

1) Methodology

Sparseness level is applied by setting a number of learning parameter to zero. It is noted from Equation (2) that if \((\alpha_{ij} \land \beta_{ij}) = 0\) for a given connection, then \(w_{ij}(k+1) = w_{ij}(k)\)
or \( v_{ij}(k+1) = v_{ij}(k) \). In other words, the weight for the specific connection does not change, therefore simulating the absence of connections. For the remaining connections, the learning parameter is set slightly lower than the threshold found with (3). Both learning and recall followed the same procedure as the one described in the previous section.

2) Results

Graphs 5a to 5c show that sparseness does not have an impact on the performances during recall with additive random noise, except for the 70% sparseness condition that underperformed the other conditions by an average of 24% in 5a, 24% in 5b and 28% in 5c. However, on the recall task with pixel flipped (5d to 5f), the performances of the sparse network are slightly improved as compared to the standard fully connected BHM (0% sparseness) where the 10% condition is slightly better than the other conditions in 5d and the 30% condition is better than all other conditions in 5e and 5f. The 70% sparse BHM also underperforms the standard BHM by an average of 19% in 5d, 17% in 5e and 18% in 5f. The results in graph 5a to 5f show that the level of
sparseness tolerated by the network is a function of the dimensionalities of the input and its association. For an equal level of sparseness, lower dimensionalities would lead to reduced performances and higher dimensionalities to increased performances. For example, for a sparseness level of 70% and randomly distributed noise of 1.5 level (graph 5a to 5c), the performances are of 0.74, 0.87 and 0.92 for respective dimensionalities of 100, 300 and 500. We expected that with higher dimensionalities, a 70% sparse BHM could bear similar performances than a standard BHM.

The performances of the BHM for varying levels of memory load, using a network of 300 units, show fairly different results. As would be expected, the performances in graphs 5g to 5i decrease as memory load increases for every condition of sparseness. However, the 70% and 50% sparse BHM conditions are much more affected by increased memory loads as compared to the standard BHM. When the memory load was set to 15%, the 70% sparse BHM underperformed the standard BHM by an average of 26% on the recall task with randomly distributed noise (5i) and by 18% on the recall task with pixel flipped (5k). When the memory load was set to 30%, the 50% sparse BHM underperformed the standard BHM by an average of 33% on the recall task with randomly distributed noise (5j) and by 20% on the recall task with pixel flipped (5l). The 70% sparse BHM could not recall the association when the memory load was set to 30%. Nevertheless, the other sparse conditions exhibited performances that were fairly similar to the standard BHM.

IV. CONCLUSION

As the results showed, the bidirectional heteroassociative memory exhibits great performances on recall tasks under noise even when spreading activation or connection sparseness is taken into account. Overall, the performances of the modified BHM with spreading activation remained very similar to that of the standard BHM, except in very particular conditions. The results are similar to those obtained in other connectionist models, such as feed-forward networks [16] and word processing neural networks [17]. In those studies it was also found that spreading activation does not affect the convergence of the network. Future research could consist of evaluating the performances of the network with spreading activation while keeping the number of learning cycles constant. In such cases, it is likely the performances would decrease as the variability is increased.

As for the influence of sparseness levels, the results are, to a certain extent, similar to the ones reported in other types of connectionist models: sparse connectivity still leads to stable solutions [26], [27], [28], [29], [32]. The network is uninfluenced by sparseness for levels up to 30%, and in right conditions, for higher levels of sparseness. It seems that past a given level of sparseness, the network lacks the connections to correctly associate the patterns, resulting in reduced performances. Future research should look into whether pruning techniques would allow for the increase of the level of sparseness in the network while maintaining optimal recall performances. Moreover, future research could investigate if the ratio of spurious attractors is independent to the level of sparseness.

By showing that the BHM model can tolerate interestingly high levels of variability, this study showed that the BHM is well suited to take into account more realistic settings. Results showed that even under variability, the performances of the BAM remain unaffected. This study suggests that the BAM is more biologically plausible as it can encompass realistic settings seen in the brain, demonstrating promising cognitive properties.

REFERENCES