

Increasing Accuracy in a Bidirectional Associative Memory through Expanded Databases

Melissa Johnson, Sylvain Chartier

University of Ottawa, Ottawa, Canada

{mjohn140, sylvain.chartier}@uottawa.ca

Abstract. Neural networks are often used in recall problems when there is noisy input and many sophisticated algorithms have been designed to help the recall process. Most cases use either learning rule adjustments, or more recently prototype learning. The question remains though of how to handle cases where there are multiple representations (exemplars) of a pattern. This paper evaluates three types of association methods: circular association method where the exemplars form a loop, linear association method where the exemplars are linked together forming a line ending in the master template, and many-to-one association method where all the exemplars point to the master template. The question asked is if using these exemplars benefits accuracy in noisy recall and does the association method matter. All three association methods had greater accuracy than the standard BAM recall. Overall, the many-to-one method had the greatest accuracy and was the most robust to changes in the exemplars. The accuracy of the circular association pattern is influenced by the amount of differences in the exemplars with accuracy increasing as the exemplars become increasingly different from each other. The linear association method is the least robust, and is affected by both the number of exemplars and differences in exemplars.

Keywords: Neural networks, BAM, Auto-association, Hetero-association, Learning, Recall, Exemplars, Classification

1 Introduction

Any intelligent system must be able to take in data and be able to process that data for storage or use. Unfortunately not all data is clear and concise, therefore a key requirement for any intelligent system is to be able to handle noisy or degraded data (Voss, 2007). Being able to recognize and recall noisy or degraded patterns is something that humans can do quickly but is still difficult for computers and artificial intelligence models. Currently, artificial neural networks are being used for this noisy recall problem because of their ability to develop attractors for each pattern and because of their stability and adaptability with regard to noise and pattern degradation.

More precisely, bidirectional associative memories (BAMs; Kosko, 1988) are used in order to associate two sets of patterns. Over the years, several variants of BAM models have been proposed to overcome the original model's limited storage capaci-

ties and improve its noise sensitivity (Shen & Cruz Jr, 2005). Nowadays, BAM models can store and recall all the patterns in a learning set, are robust to noise, and are able to perform pattern completion. This is the outcome of numerous sophisticated approaches that modify the learning and transmission functions (for a review, see Acevedo-Mosqueda, Yanez-Marquez & Acevedo-Mosqueda; 2013). More recently Chartier & Boukadoum (2011) proposed a BAM that uses the nonlinear feedback from a novel output function to learn online to iteratively develop weight connections that converge to a stable solution. The proposed BAM learns by only using covariance matrices, and it is among the few models that can create real-valued attractors without preprocessing. It is also able to reduce the number of spurious attractors while maintaining performance in terms of noise degradation and storage capacity. However, in all cases the learning of a given category is achieved by prototype associations. Although this is suitable for simple cases, in many situations there will be more than one representation (exemplars) that need to be associated with a given category. Therefore, the question of how multiple representations should be associated together remains. In this paper we propose to explore four types of associations and their impact on learning time and recall performance. More precisely, we will test the network on auto-association method (standard BAM), circular association method, linear association method, and many-to-one association method. Explanation of those association methods are discussed in the simulation section.

The organization of this paper consists of a description of the model: architecture, transmission, and learning, followed by the simulations and then a discussion of the results and conclusions.

2 Model

2.1 Architecture

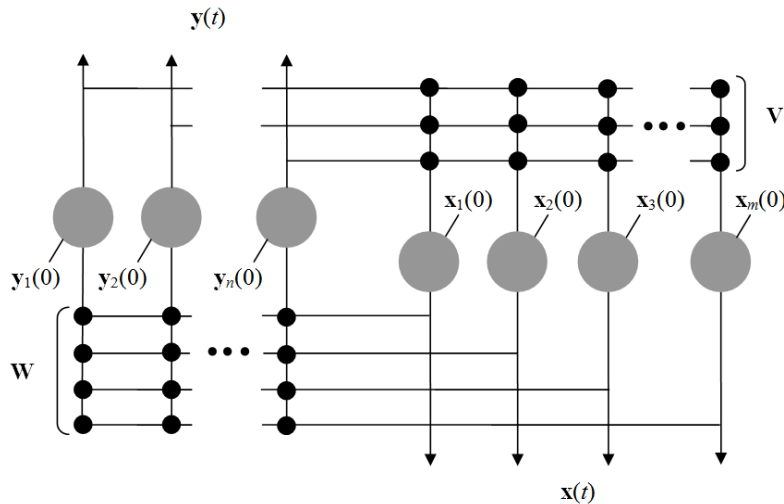


Fig. 1. BAM Network Architecture

As illustrated in Figure 1, because BAM is bi-directional, there are two initial input states (stimuli), $\mathbf{x}(0)$ and $\mathbf{y}(0)$ and \mathbf{W} and \mathbf{V} are their respective weight matrices. In the illustration, t represents the number of iterations over the network. The network is composed of two interconnected layers through which information is processed bidirectionally; the x -layer returns information to the y -layer and vice versa. The BAM neural network can be both an auto-associative and hetero-associative memory.

2.2 Transmission Function

The transmission function is based on the classic Verhulst equation extended to a cubic form with saturating limit at ± 1 (Chartier, Renaud, & Boukadoum, 2008). The transmission functions are defined by the following two equations:

$$\begin{aligned} \forall i, \dots, N, \mathbf{y}_i(t+1) &= f(\mathbf{W}\mathbf{x}_i(t)) \\ &= \begin{cases} 1 & \text{if } \mathbf{W}\mathbf{x}_i(t) > 1 \\ -1 & \text{if } \mathbf{W}\mathbf{x}_i(t) < -1 \\ (\delta + 1)\mathbf{W}\mathbf{x}_i(t) - \delta\mathbf{W}\mathbf{x}_i^3(t) & \text{else} \end{cases} \end{aligned} \quad (1)$$

$$\begin{aligned} \forall i, \dots, M, \mathbf{x}_i(t+1) &= f(\mathbf{V}\mathbf{y}_i(t)) \\ &= \begin{cases} 1 & \text{if } \mathbf{V}\mathbf{y}_i(t) > 1 \\ -1 & \text{if } \mathbf{V}\mathbf{y}_i(t) < -1 \\ (\delta + 1)\mathbf{V}\mathbf{y}_i(t) - \delta\mathbf{V}\mathbf{y}_i^3(t) & \text{else} \end{cases} \end{aligned} \quad (2)$$

where N and M are the number of units in each layer. The parameter i is the index of the respective elements during training or recall. At iteration time t , the layer contents are represented by $\mathbf{x}(t)$ and $\mathbf{y}(t)$. The weight matrices are \mathbf{W} and \mathbf{V} and the δ is the general transmission parameter. The general transmission parameter needs to be fixed at a value between 0 and 0.5 to assure fixed-point behaviour (Chartier, Renaud, & Boukadoum, 2008). This transmission function is used because it has no asymptotic behaviour when δ is between 0 and 0.5 and is therefore useable during the learning and recall. A saturating limit at the two attractors, -1 and 1 allows it to be comparable to a sigmoid type function.

2.3 Learning Rule

Most BAM models learn using a Hebbian type learning (Chartier & Boukadoum, 2011). In this model, the learning rule is expressed by the following equations:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta(\mathbf{y}(0) - \mathbf{y}(t))(\mathbf{x}(0) + \mathbf{x}(t))^T \quad (3)$$

$$\mathbf{V}(k+1) = \mathbf{V}(k) + \eta(\mathbf{x}(0) - \mathbf{x}(t))(\mathbf{y}(0) + \mathbf{y}(t))^T \quad (4)$$

where η represents the learning parameter, T is the transpose operator, and k is the learning trial. The initial inputs are $\mathbf{x}(0)$ and $\mathbf{y}(0)$ while $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are the state vectors after t iterations through the network. This learning rule can be simplified to the following equation in the case of auto association $\mathbf{y}(0)=\mathbf{x}(0)$.

$$\mathbf{W}(k + 1) = \mathbf{W}(k) + \eta(\mathbf{x}(0)\mathbf{x}^T(0) - \mathbf{x}(t)\mathbf{x}^T(t)) \quad (5)$$

$$\mathbf{V}(k + 1) = \mathbf{V}(k) + \eta(\mathbf{y}(0)\mathbf{y}^T(0) - \mathbf{y}(t)\mathbf{y}^T(t)) \quad (6)$$

Based on equations (3) and (4) the weights can only converge when $\mathbf{y}(t)=\mathbf{y}(0)$ and $\mathbf{x}(t)=\mathbf{x}(0)$. Therefore, the learning rule is linked to the network's output. In order for the association to be stored as a fixed point, η must be set according to the following condition (Chartier, Renaud, & Boukadoum, 2008):

$$\eta < \frac{1}{2(1 - 2\delta)\text{Max}[N, M]}; \quad \delta \neq 1/2$$

3 Simulation

The goal is to evaluate if expanding the dataset to include some exemplars will lead to more accurate recall of noisy stimuli, and if the type of associations between the exemplars have an impact on the accuracy of recall.

3.1 Methodology

The templates used are 7x7 pixel images of alphabetical characters (see Figure 2 for an example) flattened into a vector. Each pixel is translated into values of either 1 if the pixel is black or -1 if the pixel is white. Exemplars are created from the original template by randomly switching some (2%, 6%, 8%, or 14%) of the pixels in the templates so that values of 1 became -1 and vice versa. The creation of the noisy recall item is done by randomly flipping 30% of original templates pixels. Pixel flip at 30% is quite noisy and has been a problem for earlier versions of the BAM model (Chartier & Boukadoum, 2006). With this level of noise no recall should hit floor or ceiling (0% or 100%) during the simulations and therefore shows a full range of difference scores.

The four association methods that are being investigated can be seen in Figure 2. In the auto-association method, no exemplars are used and each stimulus is associated with itself (Figure 2A). In this case extra templates are used to keep the memory load of the neural network balanced. Memory load is typically calculated based on the dimensional space used compared to the number of templates the system is requires to learn. For all simulations in this paper, there is a 49-dimensional space capacity but the number of templates varies. By keeping the number of templates consistent for all memory association methods the memory load is consistent among them. This is shown in Figure 2 where auto-association (Figure 2A) contains 6 letters (templates) while the other associations contain 2 letters and 2 templates, making 3 patterns per category, or 6 patterns in total. The circular association (Figure 2B) has the templates and exemplars forming a circle of hetero-associations such that the last exemplar is associated with the second last, which is associated with the third last, and so on

until the first exemplar is associated with the template, the template is then associated with the last exemplar, thereby forming a complete circle. The linear association (Figure 2C) is similar to the circular association except instead of having the template associate with the last exemplar, the template is associated with itself (auto-association) forming a closed line. The final association is a many-to-one association (Figure 2D) where all exemplars are associated with the template and the template is associated with itself. In all association methods, if there are no exemplars then the methods become auto-associative.

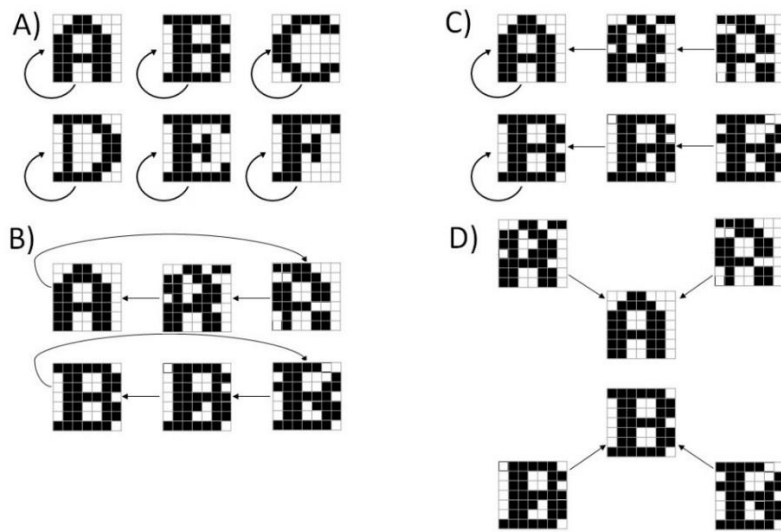


Fig. 2. Examples of templates, exemplars and their associations. A) Auto association; each item is associated with itself, B) Circular association; hetero-associations that form a loop, C) Linear association; hetero-association to one master template that is auto-associative, and D) Many-to-one association; all exemplars point to the master template

All simulations had the same variables except for the number of templates, the number of exemplars, and the level of noise in the exemplars. Number of templates plus exemplars never went above 49% memory load (24/49) of the network. All simulations were performed 150-300 times to account for the randomization of the pixel flips in both the exemplars and recall pattern. Testing of recall was done only on the template patterns that are common to all the associations. For example, in the case of Figure 2, the letters A and B would be tested while all other possible templates and exemplars are ignored.

Learning.

During the learning phase, all templates and examples are learned with their associations based on the association method being simulated. The same templates and exemplars are used for all the different association methods with the exception of the control condition which does not use the exemplars but uses extra templates to control

for memory load. The range of noise (pixel flips) in the exemplars is 2%, 6%, 8% or 14% (1, 3, 4, or 7 pixels flipped respectively) and the learning parameter, η , is set to 0.01 which respects the condition of equation 7. The transmission parameter, δ , is set to 0.1, the number of iteration before a weigh update, t , is set to 0.1 and the learning is concluded when the weights have converged.

Recall.

During recall, a noisy input pattern is the original template with 30% of the pixels flipped (15 pixels flipped). In the case of these simulations, this means that the recall pattern is always noisier than any of the exemplars used since maximum noise for the exemplars is 14%. All associations use the same recall patterns and all templates are tested. In other words, if 5 templates are originally used, no matter how many exemplars are used, 5 recall tests are performed, once per template. The transmission parameter, δ , is set to 0.1. The recall process will stop if there is no change in the output from one iteration to another or if the number of iterations have reached a maximum of 200 cycles. Recall is considered successful if the output matches the template or any of the exemplars.

3.2 Results.

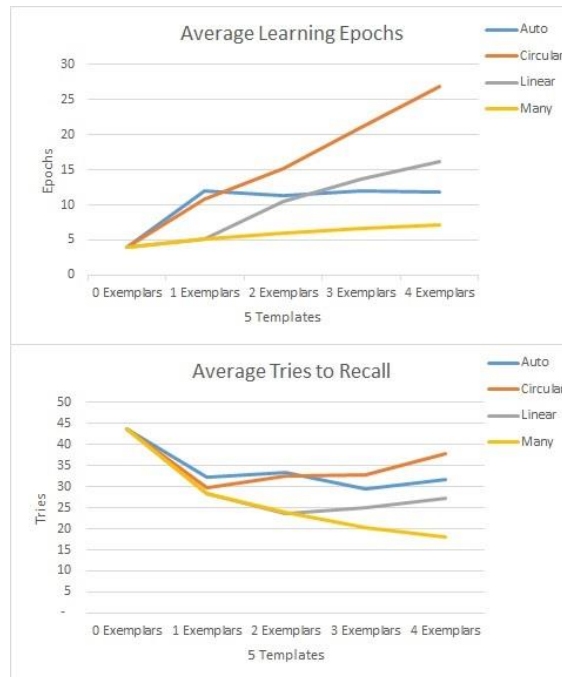


Fig. 3. On the left, the average epochs needed for learning in the different associations patterns. On the right, the amount of trials needed on average for recall. In both graphs, exemplars were set to 14% of noise, 5 templates are used with 0 to 4 exemplars per template.

Figure 3 shows that all associations have a fairly steady increase as more patterns are added. The auto-association and the many-to-one association use fewer epochs than circular and linear patterns while learning. The circular association takes the longest to learn but still consistently reaches a fixed point. Despite the added time to learn, the number of recall cycles slowly increases as more exemplars are added but they are all also still relatively close together even after 4 exemplars. The many-to-one method takes the least time to learn and the least time to recall while the normal recall, or auto-association method, is in the middle of the results. Because epochs to learn and recall cycles are all quite similar in the different association methods, it is logical to check accuracy of the methods next to see if that can differentiate the methods.

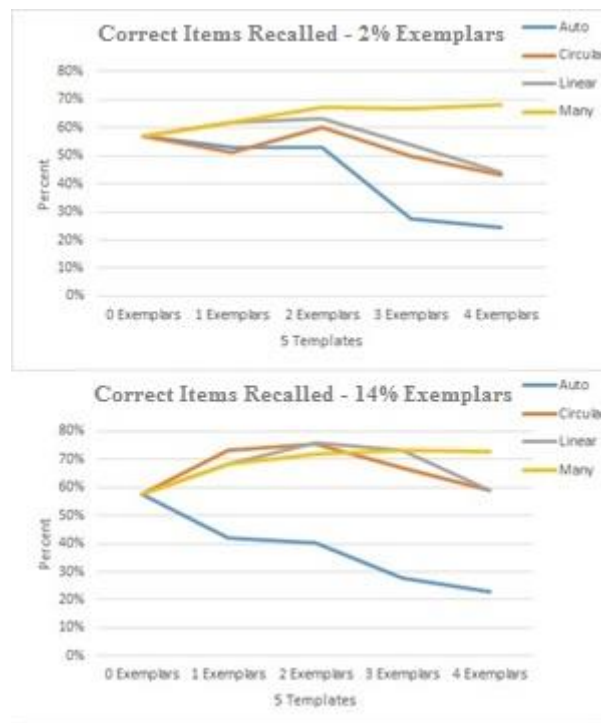


Fig. 4. Number of items correctly recalled for each of the association methods using 5 templates and 0 to 4 exemplars. Exemplars either contain 2% or 14% pixel flip (1 or 7 pixels). The stimulus to be recalled contains 30% pixel flip (15 pixels)

Figure 4 clearly shows that accuracy is improved using exemplars. If a distance of 1 pixel (2%) is used between exemplars during learning, the performance shows a slow change in performance as more exemplars are added; the auto-association method slowly reduces accuracy to end up being the least accurate while the many-to-one association method slowly increases in accuracy until it is the top performer. The difference in accuracy is more noticeable when the learning is accomplished using a

distance of 14% (7 pixel flips) between the various exemplars then using a distance of 2%. With the 14% pixel flips, all the methods increase in performance except for the auto-association which again slowly loses accuracy as more templates are added. The best performance is achieved again by multi-associations.

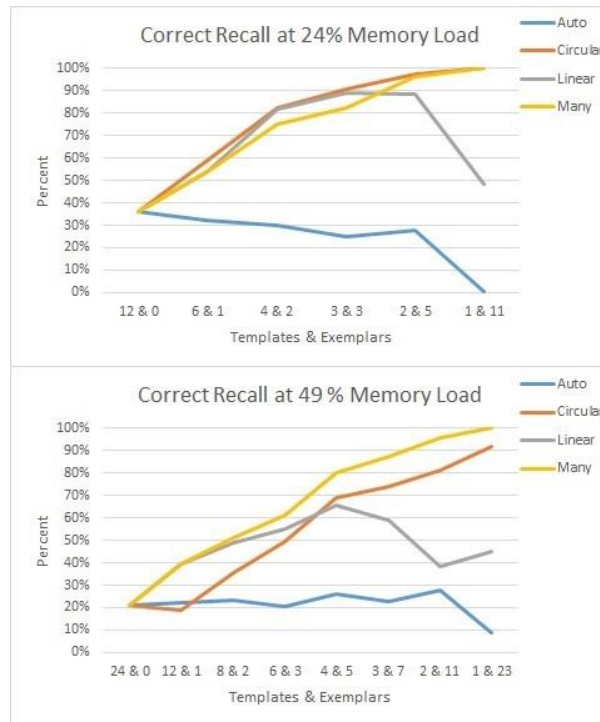


Fig. 5. Probability of correct recall while holding load steady. On the left, load is at 24% (12 patterns) and on the right load is at 49% (24 patterns). Exemplars contain 7 flipped pixels.

In the previous simulations, memory load increases as exemplars are added. Therefore we also review changes in accuracy when holding memory load steady to confirm that the increases in performance are not just because of changes in memory load. Even holding the load steady by adjusting the number of templates and exemplars, it is still clear that the new association methods outperform recall (auto) in all cases (Figure 5). This occurs both with an easy 24% memory load and a maximum efficient memory load of 49%.

The circular and many-to-one associations appear to be similar to each other and both have increasing accuracy with increasing exemplars. The linear association though appears to reach a maximum accuracy after a few exemplars are added, adding more exemplars then appears to decrease performance.

4 Discussion

Overall, the results show that expanding the dataset by including exemplars help with classification; even with minimal exemplar differences (a single pixel). The extra time it takes to learn an item is small between the different methods. Meanwhile the differences in recall accuracy appear to be quite substantial between the exemplar methods and the non-exemplar method. For example, previous research, Chartier & Boukodoum (2006) only obtains 20% accuracy with a 50% memory load using hetero-associations. Our exemplar methods which also use hetero-associations have much greater accuracy. The linear association method, the worst for performance, is closer to 50% accuracy while there is almost 100% for both the many-to-one and the circular association methods. This increase is quite impressive and a good indication that using exemplars improves recall.

While the three exemplar association methods outperform the auto association method in accuracy, the exemplar association methods are not all equal. It appears that the number of exemplars and type of exemplars matter in determining the recall accuracy of the exemplar association method.

In general, the many-to-one association method appears to be the most robust of the methods. It consistently learns the quickest and has the fewest attempts at recall before an item is found. This may be because the associations all have a similar base; all exemplars are associated with the template. This may be quicker and easy for the BAM to perform than the associations used in the other methods. The many-to-one method is also the top performer when exemplars are similar to each other and competitive when the exemplars become dissimilar. In both cases, there is a near linear increase in accuracy as exemplars are added. This shows the consistency of this method; the types of exemplars have less influence on over recall accuracy than number of exemplars.

The circular association is very susceptible to changes in the exemplars; as the difference in exemplars increase between each other and the template, accuracy increases. This suggests that there might be an ideal pattern difference that optimizes the circular association. If this is the case, the circular association method may outperform the many-to-one association method.

It is unlikely that the linear association method can compete in accuracy and usability of the circular association method or the many-to-one association method. The linear association method is affected by both the number of exemplars and the changes in the exemplars which makes it comparatively unstable and hard to optimize for best results. Conversely, the many-to-one association method is very robust and the best choice if optimization of exemplars is not an option. Due to the limited nature of the exemplars in the present simulations, it is unknown if circular association method can outperform the many-to-one association method when the exemplars are optimized.

The present study is a small study using a simple 7x7 template of alphabetical letters. This is not a realistic test to mimic human's ability to recognize degraded objects, but is an excellent starting point. Future simulations need to be run using larger templates and stricter rules in noise degradation to confirm that multi-to-one associa-

tions perform the best or to see if it is possible that a circular association would outperform depending on the type of noise introduced.

Considering the issue of the type of noise, it is known that choosing training data for classification is very important to yielding good results (Mazurowski, et al., 2008) therefore another future study could attempt to find out if there's an ideal difference between template, exemplars, and recall input pattern that maximizes accuracy for the association methods and if using this ideal case would make circular associations more accurate than the many-to-one association method.

Any and all exemplars that are created are used in this study, regardless of if the BAM actually needs them. It is possible that too many or the wrong kind of exemplars hinder performance. Therefore another possible study would be to train on exemplars that fail being recalled and not train on exemplars that can be successfully recalled. In other words, by including reinforcement learning, it may be possible to find a near ideal set of exemplars to maximize accuracy.

Acknowledgements. This research was partly supported by the Natural Sciences and Engineering Research Council of Canada

5 References

1. Abdi, H. (1988). A generalized approach for connectionist auto-associative memories: interpretation, implications and illustration for face processing. *Artificial intelligence and cognitive sciences*, 149-164.
2. Acevedo-Mosqueda, M. E., Yáñez-Márquez, C., & Acevedo-Mosqueda, M. A. (2013). Bidirectional Associative Memories: Different Approaches. *ACM Computing Surveys*, 45(2).
3. Chartier, S., & Boukadoum, M. (2006). A Bidirectional Heteroassociative Memory for Binary and Grey-Level Patterns. *IEEE Transactions on Neural Networks*, 17(2), 385-396.
4. Chartier, S., & Boukadoum, M. (2011). Encoding Static and Temporal Patterns with a Bidirectional Heteroassociative Memory. *Journal of Applied Mathematics*.
5. Chartier, S., Giguère, G., & Langlois, D. (2009). A new bidirectional heteroassociative memory encompassing correlational, competitive and topological properties. *Neural Networks*, 22(Special Issue), 568-578.
6. Chartier, S., Renaud, P., & Boukadoum, M. (2008). A nonlinear dynamic artificial neural network model of memory. *New Ideas in Psychology*, 26, 252-277.
7. Cheng, B., & Titterton, D. M. (1994). Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, 9(1), 2-54.
8. Gemmeke, J. F., Virtanen, T., & Hurmalainen, A. (2011). Exemplar-based sparse representations for noise. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7), 2067-2080.
9. Koronovskii, A. A., Trubetskov, D. I., & Khramov, A. E. (2000). Population Dynamics as a Process Obeying the Nonlinear Diffusion Equation. *Doklady Earth Sciences*, 372(2), 755-758.
10. Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A., & Tourassi, G. D. (2008). Training Neural Network Classifiers for Medical Decision Making: Effects of Imbalanced Datasets on Classification Performance. *Neural Networks*, 21(2-3), 427-436.

11. Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, 7(3), 375-402.
12. Shen, D., & Cruz Jr, J. B. (2005). Encoding strategy for maximum noise tolerance. *IEEE Trans. Neural Networks*, 16(Mar), 293-300.
13. Valentine, T., & Endo, M. (1992). Towards an Exemplar Model of Face Processing: The Effects of Race and Distinctiveness. *The Quarterly Journal of Experimental Psychology*, 44(4).
14. Voss, P. (2007). Essentials of general intelligence: The direct path to artificial general intelligence. In *Artificial general intelligence* (pp. 131-157). Berlin Heidelberg: Springer.
15. Zhang, G. P. (2000). Neural Networks for Classification: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics*, 30(4), 451-462.