General Purpose Representation and Association Machine

Part 2: Biological Implications

Lei Wei
Department of Electrical and Computer Engineering
University of Central Florida
Orlando, FL 32816, USA
lei@ee.ucf.edu

***Abstract*— Using lessons learned from error control coding, and multiple areas of life science, we propose a general purpose representation and association machine (GPRAM). In this part of paper, we illustrate our methodology, four principles, and our understanding of intelligence. We then introduce hierarchical structure and reasons to be vagueness, overcompleteness, and deliberate variation. After that, we show possible features and how to explain some visual illusions. Lastly, we illustrate a possible vague computational architecture to perform quick and rough estimation for general purpose.**

Keywords-component; formatting; style; styling; insert (key words)

I. REASON TO USE OUR METHOD

Over the last 50 years, we have collected large amounts of information about the brain. Now, we need engineers who aim to build some aspects of brain to put these back into one product, i.e., selecting carefully, and then connecting dots. GPRAM is one of these trials, but our approach is unique. In front of mountains of information, we noted that this rich information provides us with a solid basis and lots of clues, but it also blinds us to the principles and underlying truths. So, we should stand at a position close enough to see a vague picture, but at a distance enough to not be buried in the details. Distance allows me avoid being influenced by current views and thinking, so we can build up a structure based on our principles.

How to deal with the flood of information in the neural science field? In the review article [1], Fox and Raichle made the following statement. “Task-related increases in neuronal metabolism are usually small (<5%) when compared with this large resting energy consumption. Therefore, most of our knowledge about brain function comes from studying a minor component of total brain activity.” In a book review, “A neurocomputational jeremiad,” appeared in the October 2009 issue of Nature Neuroscience, it pointed out that one of key ingredients missing in neuroscience researches is the architecture of bio-brains. Our GPRAM machine covers parts of brain functions which we do not understand yet. So it is a good starting point.

As trained engineers, we always believe that if a theory is correct, then we must be able to build something and gain benefits from it. Of course, it often takes time, effort and resources to build a realistic thing that we can touch and see. It is often beyond the capability of a single person, even a group of people. But, one thing that each of us can do is to think about it, i.e., using our imagination and logical reasoning to paint a design picture. When a theory is wrong, evidence against it will appear, so we modify our theory to fit in the new evidence. A GPRAM prototype sets up a realistic goal. During the process we can learn much more on how to construct a general purpose machine and on how to represent and process massive amount information efficiently.

II. FOUR PRINCIPLES

A. Thoughts lead to four principles

When the bio-system started billions of years ago, it has no idea which tasks they would encounter later. If our GPRAM does not know which tasks need to be solved, then how do we know which representation or association is good or bad? If we cannot determine which one needs to be eliminated, then we need to search over all orders, at least as many as possible. This is often impractical. The total number of permutations for 50 ``things'' is 3.04×10^{64}. Very few tools can handle this task. One way to solve the problem is to mass duplicate the machines randomly, like bio-systems. Let them spread randomly out around the world. Each represents the subset of “things” that it has experienced and searches for some possible orders. Three points need to be considered. (a) Each individual way of representation and association must be randomized, so we can avoid any blind spot due to the way we construct the machine. (b) The machines must be able to communicate with each other in order to reduce the possibility of repeating previous searches. (c) Experiences gained from ``things'' in the outside world are important to guide them to search orders [2]. But, many rules in nature could be hidden from their obvious appearances. We must build a machine that can search for orders that may be totally irrelevant or even contradict the obvious appearances.

B. Four principles

Principle 1: Split the information processing part of GPRAM into two parts: the inner and the outer parts. The outer part will engage interactions between inside and outside; the influence of the outside world is unavoidable. The inner part needs to preserve a certain degree of possibility to reject the influence of the outside world.

Principle 2: Treat each ``thing'' in the outside world as one of many samples of its representations and freely associate those representations with little or no influence from the outside world.

Principle 3: Communications between individuals are essential for an efficient and effective GPRAM design.

Principle 4: It must have *deliberate* variation capabilities in units, signals, structures and functionalities, as well as the ability to maintain its stability.

In the past, people constantly asked me the real meaning of these four principles. Instead of explaining them, why not leave some room for readers to guess them out? We have used these principles to guide us for conceptual design of GPRAM system well beyond what in this paper. Aimed with principles, we reviewed over more than 10 different fields and visited many life science research labs across broad fields over many years.

C. Our understanding of intelligence

From the evolution of life on the earth, from our behaviors in stock markets, and from four GPRAM principles, we could extend and modify Barlow's definition on intelligence ("a good guess" [3] [4]) as follows $\text{Intelligence} = \frac{1}{3}\text{question} + \frac{1}{3}\text{answer} + \frac{1}{3}\text{dissemination}$. Looking at the 5th example in Section V GPRAM starts this action at the fundamental level.

III. Literature Review of Life Science

To build an intelligent machine we first pay special attentions to what bio-brains (including kids and animals) can do naturally, not those difficult activities which need a lot of training, and not even languages.

Unlike many existing approaches in biology and brain researches, here we need to examine the biological phenomena following the GPRAM principles. Our examinations include (1) hyperacuity in human visual systems, (2) our preference of music and abstract arts, (3) our passions for hobbies, (4) our behaviors in stock markets, (5) facts and lessons learned in our life evolution on the earth. These questions helped us to guide direction and interact with life science communities.

Over the last several years, we have been working on how hyper-acuity works in human visual systems [5]-[7]. This was a testing case for us to learn the methods in life sciences. We selected a simplest task in the human visual system, the hyper-acuity capability. In [7], we showed how to construct ideal system for various realistic environments. In order to differentiate two separation distances differed by 1/10 to 1/30th of the size of photo-receptor under general moving environments, we need a massive number of templates for very small region of uncertainty. For example, 7776 templates to deal with a speed up to 2 degree/s stimulus movement with randomly selected direction and 9 by 9 min position uncertainty region. This is for two simple dots only. The human visual system can handle various hyper-acuity tasks under very dynamic environments, yet its performance closely approach to various ideal observers, which each is optimized to a particular setting. It seems the visual system has used a super approximated template which is approximated as the superposition of many ideal templates. It may tune toward a particular template smoothly and slowly when it is exposed to some artificially designed stimuli intensively.

To understand general principles and establish a broad picture, it is not sufficient to read scientific papers from journals or books. The easiest and most effective way is to walk into their labs and talk to them, often students, directly, with questions in mind. Wisdoms often stand out clearly and sharply in unexpected conversations. This is how we connect one of important dots, the role of music signals.

We had been struggling to find a signal base in our GPRAM to make association with those abstract representations that may involve in primitive planning skills in animals and protohuman beings. We were puzzling birding singing and human preference of music and dancing. One day we had a discussion on roving attention [8]. Suddenly, a clear picture merged that we could use music-like signals as a base for primitive planning and attention. Suddenly, evidences supporting this thought stood out from many places, for example, the behaviors of tits [10], other animals [11], and the deaf boy Joseph in [9]. In briefs, the better singers of tits had better winter dominance position, survived better, and had a higher individual lifetime reproductive success. "Neither winter dominance position nor song was related to size (wing length, tarso-metatarsus length, weight)" [10]. If we will build a GPRAM brain for a tit, then we have a very simple and straight-forward answer. The GPRAM tit uses singing songs to enhance its capability, i.e., use the mind-mouth-ear-mind loop to make its music-like signal base more coherent, longer strings, more variations, so it can better predict the future.

Now let us summarize other key lessons learned from bio-systems.

[*Blurring a dot in order to see it clear*, [5]-[7]] In order to achieve hyperacuity, our eye blurs a tiny dot stimulus. If our cortex uses this blurred image to work out a result far better than sharp one, then its focus does not need to be improved. Furthermore, prepared in blurred could easily handle head and eye movements. *When we use precision* *approach, we wish each part to be as optimal as possible.* Optimizing one part may miss out opportunity to discover smart ways to use non-optimizing parts for a better and efficient overall result. Versatile approach will search for many possible combinations for a better overall result.

[*Over-complete and sparse coding*: [12] [13]] Olshausen and Field showed over-complete templates lead to sparse neural activities.

[*STDP rule*: [14]] If we apply STDP rule, it leads to discourage of small loops in code structure and implement a posterior computation. We will show this in future.

[*Early stage development*:] One of key ingredients is the early development of each part.

[*Life evolution*:] Up to 2 millions of years ago, rapid environment changes have challenged all species; particularly developing better brain structure as well as size to accommodate the need for survives. Over the last three hundreds of years, it is not the natural event, but our own motivation has led us to the industry revolution. Is this due to our unpredictable and deliberately random nature in our mind, or in Mr. Soros's term, "reflexivity" acting as a perturbation force to drive us out of the equilibrium and entered to industry revolution?

Remarks: We just list a few key items here and mention the functions which are essential to our GPRAM design. Each strategy discovered by nature could have many purposes, and for each purpose, nature could discover many solutions simultaneously using the versatile approach. For example, nature could have found many methods to perform memory. We probably know very little about these methods.

IV. HIERARCHICAL STRUCTURE OF A GPRAM SYSTEM

Now, following the example system in the first part of this paper, we purpose a hierarchical structure (well beyond [2]) using error control coding and iterative decoding.

From the higher (i.e., more vague) layer (illustrated with thick lines) to the lower (i.e., more accurate) layer (illustrated in thin lines), each is implemented with different degree of vague, deliberately variant to each other, and over-complete representation and association. Very often each layer has multiple sub-layers, and the boundaries among these layers and sub-layers are often not clearly defined. The sensor layer splits into sensor neural layer, which aims to reduce the redundancy in the sensor information, simple cell layer, which focuses on local features for accurate guesses and actions, complex cell layer which focuses on global features for coarse and quick guesses and action.

During the operation time, the GPRAM take sensor inputs, decide actions, and examine outcome of actions from sensor inputs. To achieve robustness, it will stay at higher layers. To achieve fineness, it will go into lower layers. As soon as it decides to go into a next lower layer, only those sensor and action units within the control of the current layer will be activated. Others are excluded.

Consider a GPRAM sees a scene of a man with something in his hand. If the GPRAM decides that the scene is safe, then all it needs to do is to have a rough guess of what is meant by the man's hand gesture. For example, "waving his hand" means "say hello", "pointing to somewhere" means "see this", etc. Only when the GPRAM wants to examine possible threat, it will then zoom in its focuses to "what is in the man's hand?" These kinds of actions are very efficient, but could miss opportunities.

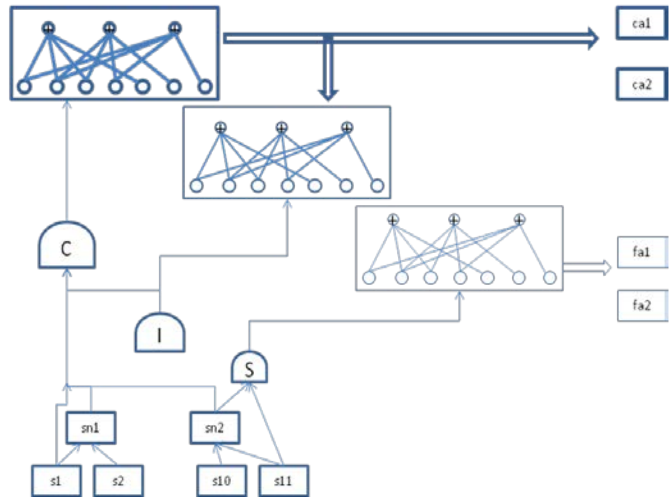

Figure 1. Illustration of hierarchical structures.

Let us define hierarchical structures in coding parts as structure. Then, we can define the "thinking" action as message-passing between different parts inside its existing structure (i.e., connections). The decision action is defined as converging to a code word. Once it has a satisfied outcome, it will significantly modify its structure. During thinking, it will slightly modify its structure a little at a time. Many methods (in AI) can be used for structure learning.

Many natural properties will be examined later on. Here, let us examine reasons to be vague using simple examples in the next section.

V. REASONS TO USE VAGUE REPRSENTATION AND ASSOCIATION

We use several simple examples to illustrate the reasons to stay at vagueness, overcompleteness, and deliberate variation in this section.

A. Illustration of vague, over-complete and variant representations and rules

It has been well known that vague representations, for example, probability measurement in Bayesian networks and a range value instead of 0 and 1 in Fuzzy logic, have pushed artificial intelligence forward. But, vagueness is mainly to reduce the complexity of representation in computation. Over-completeness, as a concept, is not efficient. Therefore we often rule it out, which can cause difficulties in implementing some good features later.

Let us see the ***first*** example. In order to measure temperature of a cup of milk with single degree accuracy from 1 to 100, we would need 100 values. But if we measure temperature with a vague measurement, say [cold: <50; hot: >50], we only need 2 values. So, the computer only needs to handle 2 values instead of 100 values; the *complexity* is reduced dramatically. However, each of us has different perspective regarding cold and hot. Furthermore, for each different task, the meaning of cold and hot could be slightly different. So, we can generate multiple sets of values: set 1 [cold: <50; hot: >50], 2 [cold: <60; hot: >60] and 3 [cold: <40; hot: >40]. This leads to vague and over-completed sets, each with slight variation. With vague, over-complete, and variant representation, we make representation much more flexible. If we do a task and find that set 1 is no good, we can try the second set, or the third. Humans are inventive. Why do we follow the rules? Why do we not create a new set (4[cold:<60; hot:>40])? Is this a case of "thinking out of a box"?

Let us see the ***second*** example. Children will try to answer almost any question we ask them. Many of the answers, which do not make sense in one instance, could make sense in another instance. If we ask a child to answer "1 plus 1=?", occasionally we get 1+1=3 for a good reason, 1 (mom)+1(dad)=3(mom+dad+me (the kid)). Does the kid think outside of the rule box? And if so, which box? If we blur the boundaries of these boxes, then the meaningless answers in one box could be meaningful in another.

Sometimes, over-complete representation can help us to reduce the *complexity* during an operation. Let us look at the ***third*** example [13]. Suppose we need to describe a dot moving along a direction on a piece of paper, what do we do? Most of us will plot an (x,y) coordinate plane, and then describe the coordinates in two real numbers. If we have 16 neurons, evenly tuning to 16 directions and each covering for 22.5 degrees roughly. Our brain will only one neuron to report the moving direction of the dot roughly, instead of two real numbers. Although it may not always be the best possible estimator, as long as upper-level processors take into account the vagueness in the message, it will work fine.

B. Illustration of vague, over-complete, variant interaction

Let us look at the ***fourth*** example. How to compute square-root of 2 using a computer? Enquiry (guess, check, refine guess): at first trial, we guess 2, then find out $2^2=4>2$; so we refine our guess to 1, and find out $1^2=1<2$, then refine our guess again, and so on. We can compute it to any accuracy.
The same idea can be used in interaction between boxes of neurons (our ***fifth*** example). Assuming two boxes (A and B) of neurons, each is well connected to do some functions, but with nothing connected in between them, how do they interact? Association links must be established. Presuming that they establish an initial link, they now can start to influence one another (say A to B). Next time, when a task is coming, it will use this association to answer and get a feedback, the second association (B to A). It uses the second association link to guide its next answer, the third association (A to B), and so on. Gradually it improves its efficiency.

Over-complete, variation, and vague in representation and association lay a very important foundation for the structure and operation in an entire GPRAM. We need to implement these at all parts. These are deliberated variation at the unit and structural levels to cover general purpose tasks, since inefficiency at one place may improve over-all efficiency, over-all flexibility, and over-all robustness.

VI. Key Features of GPRAM System and Our Explanation to Several Illusion

Dream-like experience: Let us introduce "dream"-like experiences to our GPRAM machine. Influence from heavy noise will be gradually reduced into upper layer structures since all three layers (upper structure layer, interaction layer, and sensor / operation layer) by multi-level vague hierarchies. Now, if we replay the signals in the upper layer structures, many discrepancies between the newly established influences and the existing structure can be reduced by using plasticity and querying learning methods (see our 5^{th} example). Again, we use three well-known factors in iterative decoding. (a) Message-play back can be achieved by iterative decoding, i.e., iterative computation of a posterior probabilities. (b) It can be initiated at any place using the flooding strategy. (c) It often converges to similar code words, but rarely hits the same one when the code structure becomes very large.

Decrease in sleeping time when it matures: When our system starts, its structure is not specified properly (i.e., weak codes), so switching noises can easily switch it into the sleep mode. When the system matures, the structure becomes more ridged and resilient (i.e., strong codes) so it is harder for switching noises to put the system in sleep.

Uniqueness: A substantial part of structure is determined during its initial growing phase, i.e., training with the sounds of the womb. Any future development will be heavily influenced by this structure. Even though GPRAMs will have common preferences towards music (due to similarity in the sounds of the womb and the construction procedures), each will still maintain a slight variation in its preferences.

Hierarchical efficiency: GPRAM will use hierarchical layers to zoom in and out as quickly as possible to decide on its next action.

Visual robustness: As soon as it learns at high levels to associate an object (say a hand or a face) with different tasks (rotation or movement), it has learned the same association for all similar objects (due to vagueness). As soon as two tasks of similar patterns are connected, it learns similar movement across both tasks. As soon as a synchronization signal is tricked, all patterns on the tree that were learned in the past for different tasks, different objects, different sizes, etc are ready to match up to patterns in the sensor input signals, which were generated by transforming massive parallel signals using complex cells. So, it has face recognition and CAPTCHA ability similar to human.

We used many visual illusions to guide us in the design of a GPRAM; an example is the Muller Lyer illusion. There are many explanations to causes of visual illusions. With our GPRAM architecture, the first three illusions (a, b, c in Figure 2.) are easy to understand. The last one is a little tricky. When an image is in between two vague templates, oscillation between the two templates creates a moving sensation. So, the wheel seems to spin from time to time.

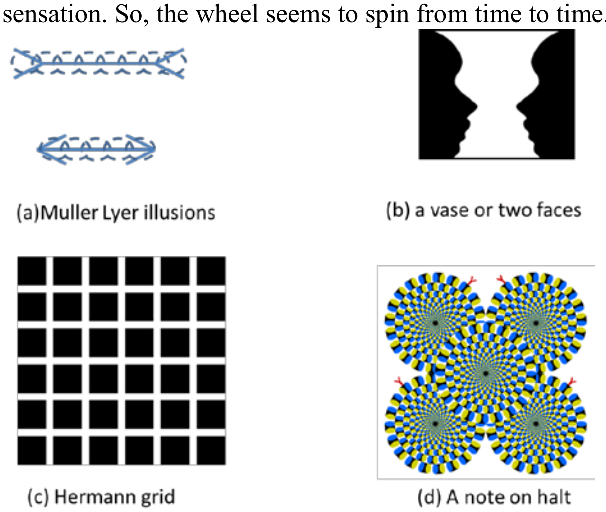

Figure 2. Four visual illusions

We leave it to the readers to explore more. These optical illusions are important since they will provide us with clues on how to construct and arrange vague templates in GPRAM.

VII. CONCLUSIONS AND DISCUSSION

In this part paper, we aim to use simple examples and illustrations to open up readers' mind for our GPRAM system in biological aspect. Combining with the first part of paper, we hope that readers can understand our goal, approach, and possible outcomes. In future, we will further discuss the roles of music-like signals, the roles of abstract art-like patterns, linkage between STDP and coding, and coding prospects of GPRAM systems, etc. All these efforts will lead to unique computer architecture to do vague computation and one step closer toward singularity [15]. Here we use the last simple example to illustrate how a vague computer works.

How do you represent vague operation? How do you describe hierarchy in music signals using vague measurement and operation? Let us use piano keys to vaguely describe the mathematical operation and location arrangement of two patches in an image. We can describe two expressions, (a) x+y=z, and (b) patch x is at the left-hand side of patch y by z inches, using one abstract form (middle A key for x, middle C key for y, the gap between for z). During its training, a GPRAM finds a code structure which links many fundamental elements together. During operations, the GPRAM asks each fundamental element to do some kinds of rough estimations for a particular aspect of the task (asking step in our intelligent model and also the 5th example). The estimated value can then be tuned up and down, just like a musician's fingers move along piano keys. These estimated pieces are then sent (answering step) and passed along the code structure to other elements (disseminating step) to decide whether it needs to tune next estimated value up or down. The GPRAM does this several times until it finally converges on a code-word that fits all of the task constraints, and then it outputs a possible solution.

REFERENCES

[1] M. D. Fox and M. E. Raichle, ``Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," Nature Reviews Neuroscience 8, 700-711 (September 2007).

[2] J. Hawkins, *On Intelligence*, ACM, New York, USA, 2004.

[3] H.B. Barlow , Intelligence, guesswork, language, Nature 304 (1983) 207-210

[4] Barlow, H.B. 1987. Intelligence: The art of good guesswork. In The Oxford Companion to the Mind, R.L. Gregory, ed., pp. 381-383. Oxford University Press, Oxford

[5] G. Westheimer, ``Visual acuity and hyperacuity, " Invest. Ophthalmol. 14, 570-572 (1975).

[6] S. A. Klein and D. M. Levi, ``Hyperacuity thresholds of 1 sec: theoretical predictions and empirical validation, " J. Opt. Soc. Am. 75, 1170-1190 (1985).

[7] L. Wei, D. M. Levi, R. Li, S. Klein, ``Feasibility Study on a hyperacuity Device with Motion Uncertainty: Two-point Stimuli," *IEEE Trans. on Systems. Man, and Cybernetics*, Apr. 2007, pp. 385-97.

[8] J.Y. Zhang, S. G. Kuai, L. Q. Xiao, S. A. Klein, D. M. Levi, C. Yu, "Stimulus Coding Rules for Perceptual Learning," PLoS Biol. 2008 August; 6(8): e197.

[9] W. H. Calvin, *How Brains Think*, BasicBooks, HarperCollins Publishers, Inc, 1996.

[10] M. Lambrechts and A.A. Dhondt, "Male quality, reproduction, and survival in the great tit (Parus major)," *Behav. Ecol. Sociobiol.* 1986, 19:57-63.

[11] A. Zahavi and A. Zahavi, *The Handicap Princile, A missing Piece of Darwin's Puzzle*, Oxford University Press, New York, 1997.

[12] Olshausen BA, Field DJ. (1996). ``Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images." *Nature*, 381: 607-609

[13] B. A. Olshausen and D. J. Field: Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision Res., 37(23) (1997).

[14] Froemke, R.C. and Dan, Y. (2002). ``Spike timing-dependent synaptic modification induced by natural spike trains." *Nature*, 416, 433-438.

[15] Ray Kurzweil, *The Singularity is Near,* The Penguin Group, USA, 2005.