

Optical Neural Network and Education

Francis T.S. Yu
Department of Electrical Engineering
The Pennsylvania State University
University Park, PA 16802
E-mail: fty1@psu.edu

ABSTRACT

As we know about it, the more we try to learn, the more we cannot learn. The thought of learning is quite different from college to college, and from person to person. In this paper, I shall discuss some of the major differences in the thoughts of learning (or rather education), as from the point of view of artificial neural network. By the virtues of understanding a neural network learning, we would be able to educate our students more efficiently and effectively. Particularly with the current rapid-trend of information science and technology, we may have to force ourselves to educate our students differently. Experimental results, as derived from some of our optical neural network models, as related to education, will be demonstrated and interpreted.

I. INTRODUCTION

Before I get started, there is a question to be asked, what is the motivation of writing this article? The apparent answer is that, to cope with the rapid changes of sciences and technology, we have to educate our science and engineering students differently. The fact is that our brains or rather the neural network (NN) that we inherited is not evolved (or designed) to behave like a computer, otherwise, we would not have developed the computers. Actually, the more you try to remember, the more you cannot remember! Even if we force ourselves to remember (or memorize) a lot of materials, most of them were quickly forgotten! For example, if one forces himself to memorize the detail content of a course, the question is how much the memorization that he can actually remember, as time passes by. From this simple experience, we know that there is close relationship between the way we learn and the usage of our NNs (i.e., brains).

In this article, I will briefly discuss the basic distinctions of a personal computer (PC) and a neural network(NN). Samples of some of the NN operation as related to educational process will be give in the form of optical simulations.

II. DIFFERENCES BETWEEN NNs AND PCs

Digital computer usually uses one or several central processing units (CPUs), while NN consists of many-many processing elements (called neurons). Computer performs the sequential (or pseudo-parallel) processings, while NN performs high density parallel processings. Computer uses programs to process the information, but NN uses learning rules instead. Thus, we see that

computer is purposely designed for computational problems, while NN is evolved for cognitive task operations, such as understanding a language or pattern recognition. In fact these tasks are still beyond the reach of modern computers.

Nevertheless, as we know that the survival instinct of all the animals on this planet earth is much depending on the cognitive task operation, such as instant recognition. This must be one of primary reasons that our brain behaves like a neural net, instead of a computer. It is also one of the major reasons that motivated us to develop the computers, to compensate our inadequacy. On the other hand, if there are some living species on a far distant planet, and their survival instincts are depending on computation, then their brains would have evolved like computers!

III. ASPECT of NNs

A human brain consists of millions of neurons and each neuron is interconnected to thousands of other neurons. The technology underlying the development of artificial neural networks(ANNs) draws heavily on cognitive psychology and biological models of human memory, learning, and perception. They attempt to implement the useful computational properties exhibited by the biological nervous systems. In other words, NN researchers try to simulate the structure of a massively interconnected biological NN, by which information processing can be performed in a parallel and associative manner.

NN consists of a collection of processing elements called neurons. Each neuron has many input signals, but only one output signal, which is fanned out along many pathways to other neurons. These pathways connect to the outer neurons from a network, as shown in Fig. 1. The processing that each neuron does is determined by a transfer function – a mathematical formula that defines the neuron's output as a function of the input signals. Every connection as attached to a neuron has an adaptive coefficient called weight assigned to it. The weight determines the connection strength between neurons. The weights can be changed by a learning rule – a rule that modifies the weights in response to the input signals and the value supplied by the transfer function. The learning rule allows the neuron's response to change with time, depending upon the nature of the input signals. This is the means by which the network adapts itself to the environment and to organize the information within itself – in short time learning manner. There are however two basic kinds of learnings, namely *supervised* and *unsupervised*, as will be discussed later.

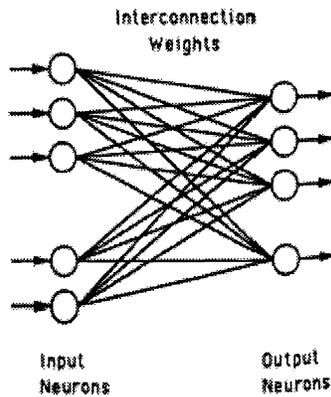


Figure 1. A single layer fully interconnected neural network.

Nevertheless, a NN of N neurons has N^2 interconnections. The transfer function of the network can be a threshold function that is, a step function, making the state of a neuron binary (0 or 1), or a sigmoid function, which gives a neuron analog value. As illustrated in Fig. 2, the state of the i th neuron in the network can be expressed mathematically by the following iterative equation also called retrieval equation

$$u_i = f \left\{ \sum_{j=1}^N T_{ij} u_j - \theta_i \right\}, \quad (1)$$

where u_i is the activation potential of the i th neuron, T_{ij} is the interconnection weight (called associative memory) between the j th neuron and the i th neuron, and f is a nonlinear transfer function, which we assume is a thresholding function that can be written as

$$f(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leq 0, \end{cases} \quad (2)$$

Thus we see that the operation of a neuron is simply a nonlinear transfer operation of the sum of weighted input signals. In other words, the mathematical expression can be represented by a matrix-vector product with a nonlinear operator:

$$U = f\{TV\}, \quad (3)$$

where U is the state vector or output vector of the neural network, V stands for the input vector, and T is defined as the interconnection weight matrix (IWM) or associative memory or simply *memory matrix*.

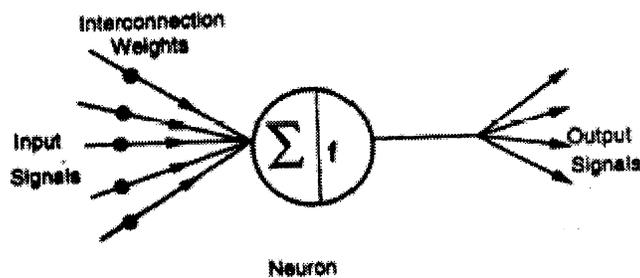


Figure 2. A mathematical model of a neuron.

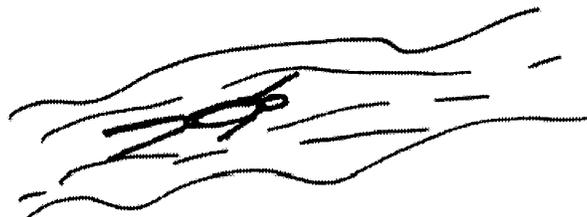
IV. ASSOCIATIVE MEMORY

They were a great deal of interest for implementing associative memory with optical techniques. Much of it is in fact centered at the implementation of NN models. It is well known brains store and retrieve information based on *process of association*. In fact, the associative memory is one of the best technique for information storage and for retrieval. Associative memory is also capable of retrieving partial or erroneous information. It is robust to noise and perturbation. And it has the ability to select the closeness and similar information.

However, a question may be raised, what is an associative memory? The answer to the question can be illustrated by the following interesting examples:

My colleagues frequently asked me why Chinese and Japanese names are very difficult to remember. The answer to this question is that. By remembering the Chinese or Japanese name is usually by the meaning by not by the sound. For example, I have met Dr Kazuo Nagakawa from Japan for a number of years, and we have met occasionally in the international conferences. It was very embarrassing for me not being able to remember his name. Until he gave me his business card with his Japanese name 中川一夫 attached, then I see the meaning of his name is that 'a gentleman is swimming in the middle of a river', as depicted in Fig. 3. Since then, I have never forgotten his name.

Japanese Name: Kazuo Nagakawa
中川一夫



A gentleman is swimming in the middle of a river

Figure 3. An example of using associative memory.

As we may know that, one can use the process of association to improve his memorization. For example, to remember a telephone number, one can associate the numerical numbers with the forms of addition, subtraction, sequencing, partition, and even the rhythm of the numbers. If one uses this associative processing, he may remember more telephone numbers than the others without using it. Nevertheless, as compared with a pocket-size computer, the number of telephone numbers that he can remember is still very limited!

V. AN OPTICAL NN

Due to the recent development of liquid crystal televisions (LCTVs), the imaging quality is very close to the commercially available high resolution video monitors. For instance, the contrast ratio can be higher than 30:1, and the dynamic range is about 16 gray level. The optical architecture that we are presenting is shown in Fig. 4. However, one question remains unanswered, why optical NN? The answer to this question is that, optics offers high density resolution, parallel processing, and massive connectivity, that prompted us to look into the possibility of optical NN implementation.

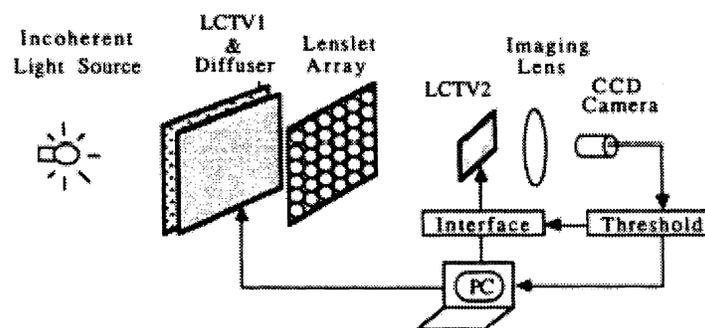


Figure 4. Schematic diagram of a compact optical ANN.

VI. NN MODELS

Although the Hopfield NN can retrieve erroneous and partial patterns, the construction of the Hopfield net is primarily based on the *intrapattern association*, which ignores the association among the stored exemplars. In other words, Hopfield model would have a limited storage capacity and it is not effective or even capable of retrieving similar patterns. One of the alternative approaches is called *interpatern association* (IPA) neural network. An example illustrating the IPA relationship is shown in Fig. 4, in which we assume that Tony and George are identical twins. Although they are very similar, we can still identify them quite easily with their special features; for examples hair and mustache. Needless to say that, by using a simple logic operation rule, an IPA NN can be constructed. As we know that, human brains are very good for recognizing the features, but it is not frequently exploited for educating our students. For example, we tend to encourage our students to memorize too many detail information, which are bound to be quickly forgotten. The fact is that the more we try to remember, the more we

cannot remember. However, if we stress more feature and conceptual learning, it will be a more efficient educational learning process.

Let us provide an experimental result as obtained from the Hopfield and the IPA NNs for illustration, as shown in Fig. 6. As we see that, the memory matrix as obtained by the IPA model has a larger storage capacity. This implies that, by using the *conceptual* or *feature learning*, one is capable of learning more information. And it is also easier to retrieve the information.

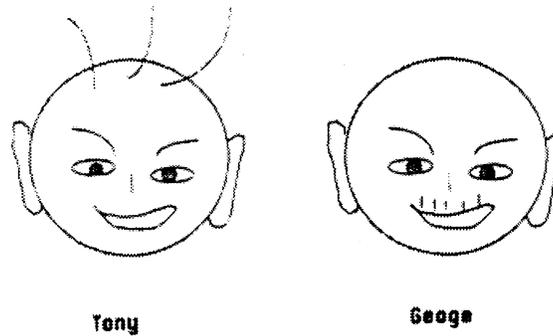


Figure 5. Concept of interpattern association (IPA).

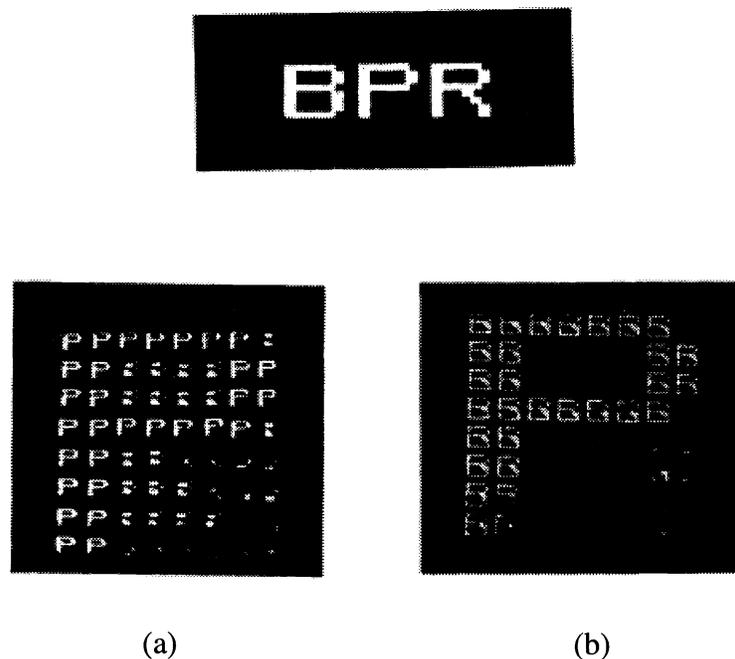


Figure 6. 8×8 memory matrices: (a) Obtained from Hopfield NN, (b) Obtained from IPA NN and BPR is the training set.

VII. UNSUPERVISED LEARNING

There are two types of learning processes; *supervised* and *unsupervised* learning. If a teacher first organizes the information and then teaches the students, then the learning process is obviously supervised (e.g., Hopfield and IPA models). On the other hand, for unsupervised learning (called *self-learning*), the students are learning by themselves based on some learning rules and their past experiences.

Since we live in the ocean of information entropy, learning is in fact a life-long process. In order to cope with the rapid changes of the technological science, we should encourage our students to educate themselves, i.e., *self-learning* or *unsupervised learning*. In reality, self-learning is a more efficient learning process, since there is no pressure applied and it is more fun. One of the inefficient ways of learning is in fact by actually taking a course. For example, in order to earn a reasonable grade, we have to force ourselves to memorize a lot of excessive and unnecessary information (e.g., equations), which are bound to be quickly forgotten.

A couple of examples as obtained from the optical standpoint will be presented in the following:

Since Kohonen model is one of the simplest self-organizing algorithms, which we will be using for our illustrations. Generally speaking, knowledge representation in human brains is at different levels of abstraction and assumes the form of a *feature map*. The Kohonen model suggests a simple learning rule by continuing adjusting the interconnection weights between input and output neurons based on the matching score between the input and the memory, a *feature map* can be constructed. By implementing the Kohonen's algorithm in the optical NN (Fig. 5) an unsupervised feature map is obtained in Fig. 7 as for interpretation. This is an interesting example similar to our young age experience. As we were memorizing the multiplication table during our grade school. We were not actually understanding the axiom of multiplication, but we have learned it.

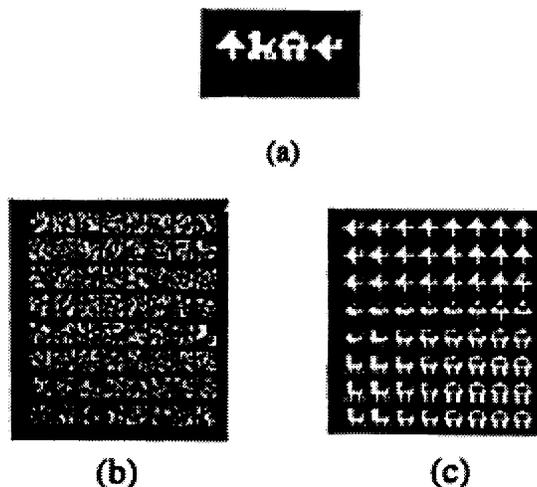


Figure 7. Demonstration of an unsupervised NN: (a) Input training sets (b) Initial memory matrix, (c) The resultant feature-map memory matrix.

Following the preceding example, we have assumed a fast and a slow learner, represented by set of linear-learning curves in Fig. 8. The constructed feature maps as observed at output plane of the optical NN, are shown in Figs. 9a and b, respectively. In these figures, we see that the slow learner has the disadvantage in speed, its feature map seems to be more topologically organized. This example is rather similar to the real life experience. As we usually see that, fast learner has the advantage of speed, but his memory is usually more disorganized and easier to forget.

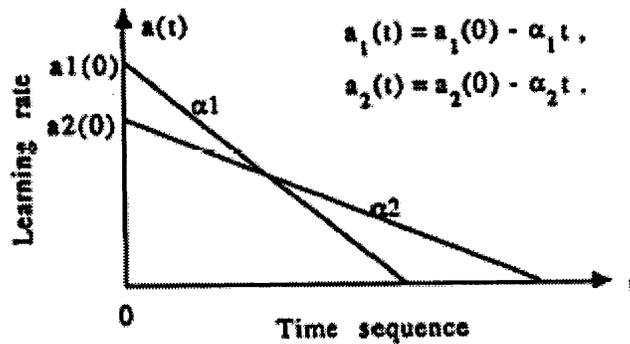


Figure 8. Fast and slow learners.

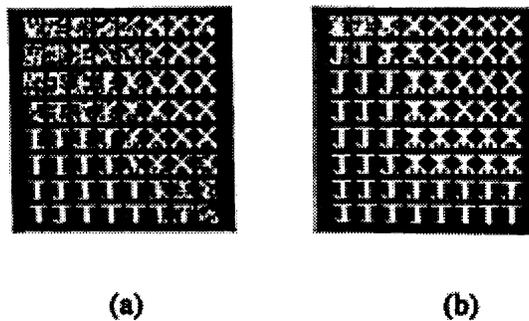


Figure 9. Memory matrix topology. (a) Obtained from the fast learner. (b) Obtained from the slow learner. [X I J T] is the training set.

VIII. LEARNING EFFICIENTLY

Efficient education depends on both *supervised* and *unsupervised* learnings. In the elementary and secondary education, the students need more supervised education. However, in college and particularly in graduate years, we should emphasize more the unsupervised learning. Nevertheless, these two learning processes cannot be mutually separated and should be coupled (or matched) properly to achieve a total quality of education. Generally speaking, good quality of education depends on good teaching and good learning.

FINAL REMARKS

A picture is worth more than a thousand words, for which we see that learning by optical means would be certainly more efficient. In view of preceding NN results as related to the learning process, the efficient education has to be from the result of efficient usages of NN (i.e., brain). In other words, as the students become more mature, this education should be more and more depending on unsupervised learnings.

Finally, to cope with the rapid changes of science and technology, we have to teach our students how to learn such that their NNs (i.e., brain) can be able to absorb the information more efficiently. For example:

- Teaches the physical significance, conceptual and philosophical content, rather than the unnecessary detail materials.
- Encourages the unsupervised learning, because it is more efficient and more fun.
- Uses the process of association to enhance the students' storage capacities.
- And others.

Once again, we remark that our brains are not designed to behave like a computer, otherwise we would not develop computers to assist us!