Some logical and numerical aspects of pattern recognition and artificial intelligence

by W. CLARK NAYLOR
IBM Corporation
Rochester, Minnesota

INTRODUCTION

Artificial Intelligence has received the attentions and contributions of workers in many varied disciplines and of many varied interests. As a result there has arisen a large and diverse body of research literature in the field. The task of sorting out and comparing some threads of continuity through this rich and variegated tapestry presents a tempting prospect.

In this article we define and compare two contrasting pattern recognition approaches. We trace their divergent paths of development from their common origin and emphasize their complementary nature. Finally, we propose their eventual reconciliation and suggest some potentially fruitful lines of development.

Threads of continuity

In 1961 Hawkins examined the state-of-the-art of self-organizing machines and traced some historical developments from early brain models to later computer implementations. This report builds on Hawkins' historical review, emphasizing two separate lines of development and extending them into more recent pattern recognition efforts.

Figure 1 displays two lines of relevant publications by author. At the origin of the lines we indicate the neuron. Section A (below) discusses some of what is known and postulated about the behavior of natural neurons. Section B follows the line of development indicated along the horizontal axis of Figure 1, and emphasizes the logical aspects of pattern recognition. In Section C we follow a line of development displayed along the vertical axis and emphasize the numerical aspects of pattern recognition. In Section D we discuss some of the problems of dealing with both aspects of the pattern classification problem at once.

A. Natural neurons

In nature, an organism interacts with its environment to enhance its chances for survival and propagation. The more an organism is capable of rapid, complex, adaptive behavior, the more effective its interaction can be. In the animal world special sensors, effectors, and associated nervous systems have developed to achieve this rapid, complex behavior. And although the complexity of the nervous systems varies greatly from the lowest to the highest animals, the properties and behavior of the basic nerve cell, the neuron, remain amazingly constant.
The neuron is a cell specialized for conducting electrical signals. It is the cell of which all nervous systems are constructed. In vertebrates, bushy dendrites extend from the cell body to receive afferent excitation and conduct it to the axon. The axon, on receiving sufficient excitation, "fires" and conducts a spike pulse along its length to the axonal branches. There the excitation is communicated across various synaptic junctions to succeeding dendrites, and so on. After firing, the cell enters a refractory state during which it rests and recharges its membranes in preparation for the next firing. Some neurons can repeat this cycle hundreds of times a second.

Many neuron configurations exist. Neurons may have long axons, very short axons, several axons, or no apparent axons at all. They may have many dendrites or no discernible dendrites. Dendrites and axons may be virtually indistinguishable. Likewise, many varieties of synapses exist. Some transmit electrically, some transmit chemically. Some transmit axon-to-axon and some transmit dendrite-to-dendrite. Probably the most interesting are the synapses between axonal branches and soma or dendrites, typical in vertebrate brain cells.

B. Logical development

A network of axons, each capable of a binary all-or-none response, is strongly suggestive of switching theory and logic, and much of the work in pattern recognition and artificial intelligence is based on this observation.

Rashevsky3 in 1938 was perhaps the first to postulate that nets of such binary axons could perform certain decision and memory functions. McCulloch and Pitts5 in 1943 formalized these concepts and showed that the behavior of such nets could be described in terms of Boolean algebra. Later, Lettvin, Maturana, McCulloch, and Pitts5 in 1959, and Verzeano and Negishi8 in 1960, were able to experimentally substantiate some of these ideas.

In 1959, Unger7 described a pattern recognition program in which he used the logical structure of a binary tree to separate an alphabet of 36 alphanumeric characters. In 1961, Kochen8 described a concept formation program which could adaptively derive its own Boolean sum of products from its experience with the data. And, in 1967, Minsky8 considered general machines composed of McCulloch-Pitts neurons. He established several universal bases for building finite automata, and showed that a very simple "refractory cell" formed such a base.

C. Numerical development

While the logical development exploits the logical ability of the axon behavior, it greatly oversimplifies or largely ignores the role of the synapse.

In 1949, Hebb9 suggested that perhaps the synapse provided the site for permanent memory. He postulated that the ability of the axonal branches and the dendrites to form graded potentials, and the ability of the synapse to differentially attenuate and integrate the influence of many impinging signals, might somehow change as a function of learning. In 1958, Rosenblatt11 incorporated these and other ideas into a model he called the Perceptron. At about the same time, Widrow12 began experiments with similar analog models he called Adalines. Many workers18-21 showed the ability of these models to implement linear decision surfaces, and the ability of certain training procedures to converge to feasible surfaces. In 1963, Rosen2 employed quadratic programming to obtain optimal decision surfaces for both linear and quadratic models. In 1964, Mangasarian23 obtained optimal decision surfaces using linear programming. Based on a Bayesian statistical approach, Specht,24 in 1967, derived optimal decision surfaces for general nth order polynomials.

D. Combined logical and numerical aspects

In his critical review of Artificial Intelligence work, Dreyfus29 addressed himself primarily to workers specializing in logical methods. In criticizing the assumptions of Newell, Shaw, and Simon50 he said "they do not even consider the possibility that the brain might process information in an entirely different way than a computer—that information might, for example, be processed globally the way a resistor analogue solves the problem of the minimal path through a network." In the present context, this and other similar comments in his paper seem to be suggestions for more careful consideration of numerical, as well as logical, methods.

While it appears that some workers have been applying logical tools to geometrical tasks, it also appears that other workers have been applying geometrical tools to logical tasks. For example, in the layered machines mentioned in Nilsson,25 it is necessary for the first layer of linear decision surfaces to completely partition the input space. Succeeding layers of linear decision surfaces then operate on the output of previous layers and so on. However, when the input space has been completely partitioned, it has been mapped without conflict onto the vertices of a hypercube. When this has been accomplished, only a problem in Boolean logic remains, and it seems a little wasteful
to use additional layers of linear decision surfaces for this task. Moreover, no general training procedures for such machines have yet been found.

While those workers mentioned in Section B have had success in dealing with the logical aspects of the pattern recognition problem, and those workers in Section C have had success in dealing with the numerical aspects, few workers have been successful in dealing with both aspects at once. However, some recent approaches appear very promising in this direction. In 1965, Casey described a program for reducing the dimensionality of sets of patterns. This would appear to be a good first step toward discovering the structure of a problem. Ball and Hall’s program to automatically cluster sets of patterns can be viewed as a process for finding logical structures of imbedded numerical decision surfaces. The most clear-cut example in this direction is the 1968 program of Mangasarian. This program iteratively builds a ternary tree of linear decision surfaces. Each surface is designed to be the optimum for its level on the tree, and the tree is indefinitely expandable.

The complementary nature of logic and geometry in pattern recognition

We would like to argue in the following sections that the two divergent lines of development pursued in the previous sections are not alternate approaches to the same problem but rather complementary approaches to that problem. That is, that a general approach must involve both aspects and that an approach emphasizing only one aspect must be somehow incomplete. This argument must be based on efficiency rather than ultimate effectiveness since either approach may be employed to eventually obtain a very good approximation to the desired result.

A. Set theory and pattern recognition

If we view pattern recognition in a set theoretic framework, the roles played by the two ordering relations, set membership, \( \in \), and set inclusion, \( ( \subseteq \), are very significant. If we are dealing with sets (or patterns) of real vectors, we see that we have two distinct algebras involved.

Among the sets themselves, we have the algebra of set combination or logic. Among the members of the sets, we have the algebra of real numbers, arithmetic or geometry. The difference in emphasis evident in the logical and numerical developments amounts to a difference in emphasis on the roles of the two algebras involved. Thus, in the logical development, the ultimate classes are composed of complex combinations of sets with very simple membership criteria. The algebra of set combination (viz logic) is strongly emphasized, while that holding among set members (viz geometry or arithmetic) is largely ignored. On the other hand, in the numerical development the ultimate classes have no apparent constituent sets and the criteria of set membership must bear the whole burden of the classification task. Thus the algebra holding among the set members (viz arithmetic or geometry) is strongly emphasized while the algebra among the sets is largely ignored.

B. Example

The complementary nature of the logical and numerical approaches may be likened to the complementary nature of Fourier and polynomial series. The Fourier series may be used to approximate a straight line, and the polynomial may be used to approximate a sine curve, but it is an unnatural and wasteful way to use the series. Similarly, logic may do the job of geometry or geometry may do the job of logic, but it is wasteful not to put each technique to its natural use. A simple example will illustrate that sets which are simply and naturally described in terms of logical combinations of numerically defined constituent sets may be very difficult to describe by logic or geometry alone. Consider the sets of rectangles defined as follows:

A: Circumference less than 20 units
B: Area less than 9 units
C: Area more than 4 units
D: Vertical side no shorter than 1/2 the horizontal side
E: Horizontal side no shorter than 1/2 the vertical side
F: \((B \cap C) \cup (D \cap E)) \cap A\)

The set F is simple enough. Try to describe it by geometry or logic alone! Figure 2 is a sketch of set F.

C. A proposed response surface

The idea of the complementary nature of logic and geometry is simple enough. Is it possible to quantify it and illustrate it graphically? Consider the following proposed axes:

X. Average number of members per component set
Y. Average number of component sets per pattern class

Z. Percentage of correct recognitions achieved

If we classify various pattern recognition programs as \((X, Y)\) points and plot \(Z(X, Y)\) for each program, what sort of graph would result? Obviously one contour must be \(Z(0, Y) = Z(X, 0) = 0\). If we further assume \(Z\) to be continuous and monotonic then contours such as those of Figure 3 will result. Figure 4, showing the relative paths of logical and numerical developments on such a surface, illustrates graphically the relative performances of logical and numerical methods.

Some optimality criteria

Having divided the pattern recognition methods into logical and numerical classes, we will find it useful and interesting to further subdivide the numerical class according to the optimality criteria used.

In numerical analysis, if we are trying to obtain the best fit of a line to a set of points, we generate an error vector and attempt to minimize some norm of the vector. The \(P\)-norm of a vector \(y = (y_1, y_2 \ldots, y_n)\) given by:

\[
L_P = \left[ \sum_{i=1}^{n} |y_i|^{P} \right]^{1/P}
\]

is the norm most commonly used for this purpose.

The values of \(P\) commonly used are \(P = 1\), \(P = 2\) and \(P = \infty\). For \(P = 1\), we minimize the average error. For \(P = 2\), we minimize the sum of squares error. For \(P = \infty\) we minimize the maximum error (Chebyshev criterion).

In pattern recognition we have a very similar situation. For any separating surface we generate a vector of separation distances and attempt to maximize the overall separation. In analogy with the one norm, we may attempt to maximize the average separation; in
analogy with the Chebyshev $\infty$ norm, we may attempt to maximize the minimum separation; or in analogy with intermediate norms we may similarly choose a whole spectrum of optimality criteria.

It is instructive to consider the sensitivity and stability of methods employing the two criteria on the extremes of this spectrum. On the one hand, a method to maximize the minimum separation will seek out the few "worst-case" points and work on them first. Such a worst-case method will be a local, differentiative method; it will be very sensitive to local details, but very prone to over-react to noise. On the other hand, an average-case method will be a global, integrative method. It will tend to be relatively insensitive to noise, but also insensitive to local detail.

An example will illustrate this noise and detail sensitivity. Consider the sets and separating plane of Figure 5. The plane satisfies worst-case, average-case and intuitive criteria for a good separating plane.

Consider the sets and planes of Figure 6. Here set $A_2$ has been augmented by $A_3$. As long as the minimum difference between points in $B$ and $A_2$ is larger than the minimum difference between $B$ and $A_1$, $A_2$ will have no effect on the placement of plane 1, the worst-case plane. The $A_3$ information is essentially redundant. However, plane 2, the average-case plane, will move around and react to set $A_2$ as set $A_2$ moves. It may even violate one of the sets. Intuitively we would probably choose plane 1.

Consider the sets of planes of Figure 7. Here Sets $A$ and $B$ have been augmented by noise patterns. Since these noise patterns affect the minimum distance between the sets, the worst-case plane, plane 1, will react, while the average-case plane, plane 2, does not. Here we would probably intuitively choose plane 2.

Again we can tie these items to physiological considerations. Several averaging and differentiating neurons have been observed in nature—particularly in optic nerves. Apparently their actions are carefully balanced to insure sharp resolution together with noise.
insensitivity. In pattern recognition we will similarly be forced to achieve a balance in the use of average and worst-case methods.

**Proposed model for future development**

From the considerations of this report, it seems clear that a general pattern recognition device will have to perform numeric calculations carefully balanced between global integrative techniques and local differing techniques. The results of these calculations will then be combined logically to determine the result of the entire device. It also seems clear that natural nervous systems will provide an existence proof and guide in the construction of feasible pattern recognition models.

Figure 8 represents a possible logical-numerical net. At each node a numeric calculation is performed and an exiting branch is chosen. The overall branching network of nodes provides the logical structure (although a logical tree is shown for simplicity, any complex logical structure is intended).

**SUMMARY**

Pascal once said that there are two kinds of mathematical minds—logicians and mathematicians. We have indicated that there are two kinds of pattern recognition programs, logical ones and geometrical ones. In this report we have traced the historical development of these two distinct approaches. We have related them to two functions of natural nerve nets and to the two algebras of set theory. By these associations we have argued for the complementary nature of the roles played by these two aspects. In addition we have distinguished and compared the roles of average-ease and worst-case geometrical methods. Finally, for future development, we have suggested a pattern recognition model encompassing the capabilities of all these methods.

**ACKNOWLEDGMENTS**

Work leading to this report was carried out in the Computer Science Department of the University of Wisconsin and the Recognition Systems Development Department of IBM Rochester, Minnesota. The author is indebted to many people for their help and encouragement during this work: Drs. L. Uhr, J. B. Rosen, and O. L. Mangasarian at the University of Wisconsin for many stimulating and enlightening discussions; A. Hamburger, D. R. Andrews, P. H. Howard, and M. J. Kimmel of the Recognition Systems Development Department at IBM for their continued encouragement and support; and C. D. Cullum of IBM Yorktown for his helpful ideas.

**REFERENCES**

1. J K HAWKINS  
   *Self-organizing systems—A review and commentary*  
   Proc of Institute of Radio Engineers January 1961

2. T H BULLOCK A G HORRIDGE  
   *Structure and function in the nervous systems of invertebrates*  
   W H Freeman and Company New York N Y 1965

3. N RASHEVSKY  
   *Mathematical biophysics*  
   University of Chicago Press Chicago Ill 1938

4. W S McCulloch W H PITTS  
   *A logical calculus of the ideas immanent in nervous activity*  
   Bulletin of Mathematics Biophysics Vol 115 1953

5. J Y LETTVIN H R MATURANA W S McCulloch W H PITTS  
   *What the frog’s eye tells the frog’s brain*  
   Proc of Institute of Radio Engineers Vol 47 November 1959 1940–1951

6. M VERZEANO K NEGISHI  
   *Neuronal activity in cortical and thalamic networks*  
   J Gen Physiol Vol 43 supply July 1960 177–195

7. S H UNGER  
   *Pattern detection and recognition*  
   Proc of Institute of Radio Engineers October 1959

8. M KOCHEN  
   *An experimetal program for the selection of disjunctive hypothesis*  
   Proc Western J C C Vol 19 571–578 May 1961

9. M MINSKY  
   *Computation: Finite and infinite machines*  
   Prentice Hall Englewood Cliffs New Jersey 1967

10. D O HEBB  
    *Organization of behavior*  
    John Wiley and Sons Inc New York N Y 1949

11. P ROSENBLATT  
    *The perceptron—A theory of statistical separability in cognitive systems*  

---

From the collection of the Computer History Museum (www.computerhistory.org)
12 B WIDROW M E HOFF
Adaptive switching circuits

13 F ROSENBLATT
On the convergence of reinforcement procedures in simple perceptrons
Cornell Aeronautical Lab Report No VG-1196-G4 February 1960

14 R D JOSEPH
Contributions to perceptron theory
Cornell Aeronautical Lab Report No VG-1196-G7 Buffalo NY June 1960

15 H D BLOCK
The perceptron: A model for brain functioning I
Reviews of Modern Physics Vol 34 123-135 January 1962

16 A CHARNES
On some fundamental theorems of perceptron theory and their geometry
Computer and Information Sciences Spartan Books Washington D C 1964

17 A B J NOVIROFF
On convergence proofs for perceptrons
Stanford Research Institute Report Nonr 3458(00) January 1965

18 R C SINGLETON
A test for linear separability as applied to self-organizing systems—1962
Spartan Books 503-524 Washington D C 1962

19 W C RICKEY
An adaptive logic system with generalizing properties

20 T S MOTZKIN I J SCHOENBERG
The relaxation method for linear inequalities
Canadian Journal of Mathematics Vol 6 No 3 303-404 1954

21 S AGMON
The relaxation method for linear inequalities
Canadian Journal of Mathematics Vol 6 No 3 383-392 1955

22 J B ROSEN
Pattern separation by convex programming
Journal of Mathematical Analysis and Applications Vol 19 No 1 February 1965

23 O L MANGASARIAN
Linear and nonlinear separation of patterns by linear programming
Operations Research Vol 13 No 3 May 1965

24 D F SPECHT
Generation of polynomial discriminant functions for pattern recognition

25 N NILSSON
Learning machines
McGraw-Hill Inc New York N Y 1965

26 R G CASEY
Linear reduction of dimensionality in pattern recognition

27 G H BALL D J HALL
ISODATA, an iterative method of multivariate analysis and pattern classification
Proc of International Communications Conference Philadelphia June 1966

28 O L MANGASARIAN
Multi-surface method of pattern separation
(to be published)

29 H L DREYFUS
Alchemy and artificial intelligence
The RAND Corporation Santa Monica California December 1965

30 A NEWELL H H SIMON
Computer simulation of human thinking

31 I P PAVLOV
Conditioned reflexes
Oxford University Press New York N Y 1927

32 D A SCHOLL A M UTTLEY
Pattern discrimination and the visual cortex
Nature 387-388 February 28 1953

33 W A CLARK B G FARLEY
Generalizations of pattern recognition in a self-organizing system
Proc W J C C 86-91 1955

34 S B AKERS
Techniques of adaptive decision making
General Electric Company Electronics Laboratory Technical Information Series R65ELS-12 Syracuse New York October 1965

35 G L NAGY
Prospects in hyperspace: State of the art in pattern recognition

36 M D CANON C D CULLUM
The determination of optimum separating hyperplanes I. A finite step procedure

37 L UHR
Pattern recognition
John Wiley and Sons Inc New York N Y 1966

38 G S SEBESTYEN
Decision-making processes in pattern recognition
Macmillan Company New York N Y 1962