AD 673 785

THE PROBLEM OF PATTERN RECOGNITION

V. I. Varshavskii, et al

Foreign Technology Division
Wright-Patterson Air Force Base, Ohio

18 September 1967

FTD-HT-67-444

AD 673785

FOREIGN TECHNOLOGY DIVISION

THE PROBLEM OF PATTERN RECOGNITION

by

V. I. Varshavskiy and V. A. Sokolovskiy

50 YEARS

1917 1967

GOLDEN ANNIVERSARY

FOREIGN TECHNOLOGY DIVISION

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

21

THIS IS A REPRINT COPY

NOTICE

This reprint has been processed by a rapid and economical method to expedite copies to requestors who received insufficient copies or were not included in initial distribution. The rapid processing procedures used in preparing these reprints unavoidably result in low-quality pictorial illustrations. However, since the majority of requestors have been satisfied with the textual information, this type of reprint is routinely used for satisfying requests when normal extra copy stock has been exhausted.

If better quality illustrations are required, recipients should direct requests to this headquarters justifying the need for this material, and identifying the specific illustrations or photographs required by document, section, page and/or figure numbers, as appropriate.

FTD-HT- 67-444

EDITED TRANSLATION

THE PROBLEM OF PATTERN RECOGNITION

By: V. I. Varshavskiy and V . A. Sokolovskiy

English pages: 17

SOURCE: Nauchno-Tekhnicheskoye Obshchestvo Radiotekhniki i Elektrosvyazi. Kibernetika i Elektronno-Vychislitel'naya Tekhnika (Scientific and Technical Society of Radio Engineering and Electrical Communications. Cybernetics and Electronic Computer Technology). 1962, pp. 12-26.

Translated under: Contract AF33(657)-16410

TT7001718

PREPARED BY:

TRANSLATION DIVISION
FOREIGN TECHNOLOGY DIVISION
WP-AFB, OHIO.

ITIS INDEX CONTROL FORM

01 Acc Nr	68 Translation Nr	65 X Ref Acc Nr	76 Reel/Frame Nr
TT7001718	HT6700444	AR4001921	1881 1794

97 Header Clas	63 Clas	64 Control Markings	94 Expansion	40 Ctry Info
UNCL	UNCL, 0	0		UR

02 Ctry	03 Ref	04 Yr	05 Vol	06 Iss	07 B. Pg.	45 E. Pg.	10 Date
UR	0000	62	000	000	0012	0026	NONE

Transliterated Title
O PROBLEME OPOZNANIYA KONFIGURATSIY

09 English Title
THE PROBLEM OF PATTERN RECOGNITION

43 Source NAUCHNO-TEKHNICHESKOYE OBSHCHESTVO RADIOTEKHNIKI I ELEKTROSVYAZ I. KIBERNETIKA I ELEKTRONNO-VYCHISLITEL'NAYA TEKHNIKA (RUSSIAN)

42 Author
VARSHAVSKIY, V. I.

16 Co-Author
SOKOLOVSKIY, V. A.

16 Co-Author
NONE

16 Co-Author
NONE

16 Co-Author
NONE

98 Document Location

47 Subject Codes
09

39 Topic Tags: pattern recognition, learning mechanism, learning process, discrete automaton, electronic feedback

ABSTRACT : The problem of pattern recognition is examined along two directions. The first direction consists of a group of problems related to constructing reading devices consisting of comparators and a large memory. A characteristic of reading devices is the inability to change their mode of action during the operating process. This leads to a study of the second direction of the problem of pattern recognition which concerns the question of constructing learning recognition devices. Learning recognition devices are automatic machines with feedback, that is, devices which compare the responses for the input patterns with the responses for an external device ("an amplifier"). The problem of pattern recognition is also reported in terms of discrete automatic machines, that is, devices having a finite number of inputs, a finite number of outputs and a finite number of internal conditions. The automatic machine obtains information about the external medium by its sensing elements, which (in the case of living organisms they are called receptors) serve as inputs to the automatic machine. The external medium is reflected in the automatic machine through the conditions of its inputs. There is one illustration and six bibliographical references. English translation: 17 pages.

THE PROBLEM OF PATTERN RECOGNITION

V. I. Varshavskiy and

V. A. Sokolovskiy

The problem of pattern recognition, which is of great practical interest, has not as yet been formulated with sufficient clarity. We do not plan to present any methods for solving it here. Our aim will be to discuss some aspects of its formulation.

The wide range and extraordinary complexity of the problem of pattern recognition has made this work somewhat fragmentary, and the authors beg your forgiveness in advance for this shortcoming.

M. L. Tsetlin was consulted concerning the subject problem and a number of his ideas have been used here. The authors would like to take this opportunity to express their sincere appreciation to him and others whose opinions appear in this work.

In our day, computers are successfully solving a great many of the most diverse problems more rapidly and accurately than man could do. However, some aspects of man's behavior, such as a capacity for "learning" associative recall, and choosing essential variables from a context containing a large number of nonessential variables, have not as yet been reproduced with sufficient success in computers.

The successful solution of such problems depends on two factors: first, on our ability to describe the process under consideration or its analogy in computer language, and, second, on the physical limitations of computers with respect to their operating speed, memory capacity, and the equipment's input and output capacity.

If we program a chess game into a computer, we can evaluate rather clearly the quality of play simply by the number of defeated players with various abilities at the game. In the case of pattern recognition, the problem of evaluating the quality of recognition is considerably more difficult. "Pattern recognition", as a process, is defined much more poorly than a game of chess, and the concepts of "correct" or "reasonable" pattern recognition are formulated much less clearly than the conditions of winning at chess.

The problem of pattern recognition has two main directions. The first involves a group of problems relating to reader design. Readers can be automatic devices which receive visual images at the input and group them together, adducing some answer which corresponds uniquely to each group. The simplest reader, as far as logical structure is concerned, consists of a comparison unit and a large memory which stores all possible input patterns along with numbers indicating to which group the pattern belongs. When in operation, the device compares an input pattern with a table which is stored in the memory until there is a match, and then transmits to the output a number which determines membership in a certain group. It is apparent that such a device becomes quite uneconomical when there is a large number of input patterns. When such is the case, a designer turns to algorithms, which are logically more complex but provide more economical design as far as circuitry and other technical matters are concerned. As an example of a reader which has a relatively complex logical structure, we can cite Sherman's work [1] on the recognition of handwritten letters.

A fundamental characteristic of readers is the impossibility of changing their behavior while they are in operation. In some cases, the form of pattern entering the input has not been defined beforehand or can be subjected to numerous, unpredictable changes. In addition, the conditions which determine pattern classification can change from case to case. If we want a pattern recognition unit to operate under various external conditions, it is unlikely that we will have sufficient time and ability to include in it all the possible answers for all the possible situations. In order to achieve flexibility, we should make the machine itself take some responsibility for "making a decision." This leads us to the second group of problems, which pertain to recognition learning units.

Recognition learning units can be automatic devices with feedback, which are able to compare their answers to input patterns with answers from an outside unit which we call the "teacher." After being shown a series of input patterns, a recognition learning unit can change its parameters or even its internal structure in an effort to decrease the discrepancy between its answers and the "teacher's" answers. We can imagine that a recognition learning machine in each of its stages is a reader and, consequently, continues a stage until it receives a "good grade" for certain prescribed conditions from the teaching machine. The quality of a reader is determined by an evaluation which we shall mention below. It is obvious that a reader is the simplest individual case of a recognition learning unit having only one state. Therefore, we shall consider below questions connected with pattern recognition only within the framework of recognition learning units, especially as the learning process can be a convenient and, in some cases, a necessary procedure in the designing of readers.

We shall attempt to state the problem of pattern recognition in terms of automata since recognition has meaning only relative to recognition by an automatic device or an algorithm (an algorithm can be considered an automatic device).

We shall call a unit which has a finite number of inputs, a finite number of outputs, and a finite number of internal states a discrete-action automated machine. This is a common definition of an automated machine, but we must expand it somewhat. Any automated machine functions in a certain external medium and its duty is to obtain information from this medium. Automata obtains information about the surrounding environment by means of their sensors, i.e., transducers, which, in the case of living organisms, are called receptors. Receptors are the inputs of automata. In this manner, the external environment is reflected in the machine through the state of its inputs. If we establish a correspondence between one of the axes of a certain space (which we shall call the receptor-signal space) and each input of the machine, then a certain point of this space will correspond to each combination of input states. We shall call each such point a pattern. An automated machine affects in some manner the medium in which it is "submerged." This effect is not necessarily active, but it can always be assumed that the output of an automated machine is connected with the input by a feedback circuit, one of whose components is the external environment, or, at least, the environment of the feedback loops comprising the automated machine if only there is one loop comprising the external environment. If the latter condition is not fulfilled, even in implicit form, then we shall call such a unit a converter, not an automated machine. If there is a finite number of outputs, we can establish a correspondence between one of the coordinate axes of a space (which we shall call the response space) and each output of an automated machine. We shall call each point of this space a response. It is natural that patterns and responses exist in limited areas of the space for patterns and the space for responses, respectively. We shall call a pattern sequence at the input for time period t, an input event with length t. We shall call a response sequence for time period t to input event with length t, a response function with length t. We shall distinguish between machines with memory, i.e., those capable of response function with length t, where t is the memory capacity of the machine, and machines without memory, i.e., those capable of response function with length I where I represents a quantum of time or the time necessary to record one pattern. It is natural to assume that automata without memory are components of automata with memory. Let us examine some considerations relating to automata without memory, while keeping in mind that the results can be applied to automata with memory. Below we shall consider the approach to automata suggested by M. L. Tsetlin in detail in [2].

Let us propose the following description of the external medium.

For each "pattern response" pair, let us give a number which we shall call a "fine" paid by the automated machine to the environment. The amount of the fine for a given machine is some integral characteristic of the environment. The fact is that the number of inputs is limited, and the same patterns may correspond to various states of the external medium which are imperceptible by the inputs. The behavior of an automated machine in the external medium is distinguished by a tendency to minimize some function from the amount of the fine. This function obviously has to be monotonically increasing. Examples of such functions are: the mathematical expectation of the amount of the fine or the mean integral value of the fine for some finite length of time.

Let us examine the possibility of designing such a machine. We shall say that we have a set of units from which an automated machine may be constructed. Each unit has n inputs and is capable of processing some function $f(x_o \ldots x_{n-1}, d_o \ldots d_{k-1})$ from n input variables and k parameters which can change and the set of which determines the parametric state of the unit. In this manner, each unit can process, depending on the parametric state, a family of functions $\{f\}_\alpha$. If we establish a correspondence between each parameter and a certain space, which we shall call the space of the machine's parametric states, then, when the structure of the machine is fixed, each point of this space will describe a certain parametric state of the machine. For each fixed structure in the space of the machine's parametric states, the medium determines some function of the fine, which the machine strives to minimize. In this sense, the definition of an automated machine agrees with the definition of self-optimizing systems for automatic control. The number of variable parameters of the machine or the size of the space of the machine's parametric states can be so large that the usual methods for finding the minimum for a function of many variables become ineffective. The problem is aggravated by the fact that the fine function can change in time and the machine must strive not to find the absolute minimum but to maintain itself on a trajectory in the space of the machine's parametric states, on which the value of the fine function is small enough.

In our opinion, the best of the known methods is the "ravine" method proposed by I. M. Gel'fand and M. L. Tsetlin in [3]. Selfridge [4] also suggests a somewhat similar method, then refers to the method of random search as feasible, whereby a unit of the machine processes a closed set of linear functions. The machine begins to function in a random parametric state. In the initial parametric state the minimum number of parameters is determined, whose variation leads to the greatest variation in the fine function. Based on the indicated variables, which are said to be nonessential, there is a gradient slope as long as the function variation at each step of the gradient slope is large enough. When the fine function variation for one step of the gradient slope is less than some predetermined criterion, the slope ceases and the parametric state obtained is called the first minimized parametric state. After this, a new random parametric state is selected, from which there occurs a gradient slope (based on parameters which are nonessential for this state) to the second minimized parametric state. The third initial state for the gradient slope is on a straight line connecting thp first and second minimized parametric states. After the gradient slope to the third minimized parametric state, the fourth initial parametric state is on a straight line connecting the second and third minimized parametric states. Search is carried out continuously. If the fine function always remains more than a certain criterion while searching, a random choice is made for a new initial parametric state.

The indicated algorithm is particularly convenient because it is valid for operation when the fine function is time-variable, while revealing, in the process of searching, the internal organization of this function. This interpretation of the definition of an automated machine does not contradict the popular definition and is also applicable to machines with memory (in the sense indicated above) and machines with variable structure, since the structure can also be described as some set of parameters.

Now we shall endeavour to introduce some concepts concerning patterns, which will be quite useful to us below. From daily experience we know that recognition is not always effectively accomplished, i.e., we do not always relate a pattern to a certain object. Sometimes we cannot determine what a pattern means or if it means anything.

We shall divide patterns, by intuition, into meaningful and "noise" patterns. By "meaningful" we shall denote a pattern which we can identify. Apparently the most difficult task in the problem of pattern recognition is the task of recognizing meaningful patterns or the task of separating meaningful ones from "noise" patterns. In our opinion, a strict determination as to whether a pattern is meaningful or "noise" is made considerably more difficult by the fact that there is no clearly expressed boundary between them. As, for instance, from some people's point of view, all pictures by abstract painters are "noise", while others maintain they can distinguish good abstract painting from bad.

It seems reasonable to us to divide the problem of pattern recognition into two classes.

Most problems considered in well-known works on pattern recognition belong to the first class. These are problems where a refined set of patterns, consisting only of meaningful patterns, is considered in the input of the machine. In this case the machine must subdivide a set of patterns into a number of subsets called pattern forms, or must classify patterns in terms of concepts available in the machine. It is expected to give a correct answer only to those patterns which it was shown during the "learning" process, i.e., in this case, the process of generalization (which we will discuss below) is not considered. In problems of the first class, recognition is considerably simplified by very high redundancy in the input of the machine.

We shall judge from the following example the magnitude of this redundancy.

Let a machine have P input dichotomous data units; then the total number of possible patterns is 2^P, which, for the case of 332 inputs, already reaches the cosmic proportions of approximately 10^{100}. Let the patterns in the input change at a rate of 10^6 patterns/sec, i.e., at a frequency of 1 MHz; then in 100 years the machine will be able to scan approximately 10^{16} patterns, which is approximately 10^{-82}% of the total possible number of patterns. This reasoning is helpful in still another respect. We can set the upper limit for the number of patterns able to appear in the input at approximately 2^{40}, which clearly will not be surpassed. Thus, when a set of input patterns is limited to those patterns which a machine has learned, the pattern space is practically vacant and we can assign "noise" patterns arbitrary values, which, in most cases, considerably simplifies the problem.

The second class of problems involves the problem of separating meaningful patterns from "noise" patterns. We do not have here the usual problem of communication theory relating to the discrimination of a signal at a certain noise level, but, rather, the problem of discriminating patterns containing information from completely meaningless patterns. If we use communication terminology, the problem can be formulated in the following manner: along a communication channel there can be transmitted noise, a signal at a certain noise level, and a pure signal. It is necessary to determine the moment when the signal is transmitted, while knowing only that the signal differs from noise in that it possesses a certain organization. In certain specific cases the problem is solved with rather minor expenditures of equipment; however, in general, the problem is incredibly complex. It seems probable to us that this problem cannot be strictly defined at all for a general case, but must be solved in each specific case differently. In terms of automata, a "noise" pattern can obviously be defined as a pattern for which the fine function does not change when the answer does. In this case, the discrimination of "noise" patterns, it would seem, has no meaning since we assume an arbitrary response to a noise pattern. But this is not quite so. If the machine takes into account the amount of the fine derived when responding to a noise pattern, it can thereby impair the process of minimizing the fine function. Below we shall endeavor to formulate some concepts relative to both the first and second approaches.

We shall call the first approach, relating to recognition in a closed set of meaningful patterns, the task of membership classification. This task can be formulated mathematically in the following manner: it is necessary to construct in the space of receptor signals or in the pattern space a function which leads to a correspondence between each group of patterns (indicated by the "teacher") and a corresponding response. Since the grouping feature can be rather arbitrary, it is unlikely that patterns belonging to one group can be easily enough distinguished from other patterns by a simple function.

Without loss of generality, only cases of dichotomous classification can be examined. Usually, when solving the problem of dichotomous classification, we use a linear function as a separating function, i.e., we divide the pattern space into two domains by a hyperplane. The use of a hyperplane as a dividing surface is due to a number of reasons. In the first place, application of a hyperplane is the simplest, both with respect to circuitry and program; in the second place, since coordinates of a point in the pattern space, as a rule, are only zero and one, the use of a higher order function has an effect which is not commensurate with the increased complexity of application. However, the simplicity of applying a linear separation comes as no gift. Far from all pattern sets can be linearly divided. Let us consider a simple example.

Let four patterns be assigned to a rectangular receptor matrix (Figure 1). It is reasonable to relate the two upper patterns to one group and the two lower patterns to the other. In a 49-dimensional space of receptor signals the following points correspond to these patterns:

$$X_1 = (000000000111100010010001001000111 [illegible] 0000000000000),$$
$$X_2 = (0000000001111000100100010010001111000000000000000),$$
$$X_3 = (000000000111100000110000011000111 [illegible] 00000000),$$
$$X_4 = ([illegible] 0111100011000001100000111100 [illegible]).$$

It is known that two sets have a separating hyperplane only if their convex hulls do not intersect; i.e., if

$$\sum_{j=1}^{M} \lambda_j X_j \neq \sum_{i=1}^{K} \delta_i Y_i$$

when

$$\lambda_j \geqslant 0,\ \delta_i \geqslant 0,\quad \sum_{j=1}^{M} \lambda_j = \sum_{i=1}^{K} \delta_i = 1,$$

where X_j - M are points of the first set and Y_i - K are points of the second set.

The convex hull of set (X_1, X_2) is segment $t = \lambda . X_1 + (1 - \lambda) X_2$.

The convex hull of set (X_3, X_4) is segment $f = \delta X_3 + (1 - \delta) X_4 (0 \leqslant \lambda \leqslant 1$ and $0 \leqslant \delta \leqslant 1)$.

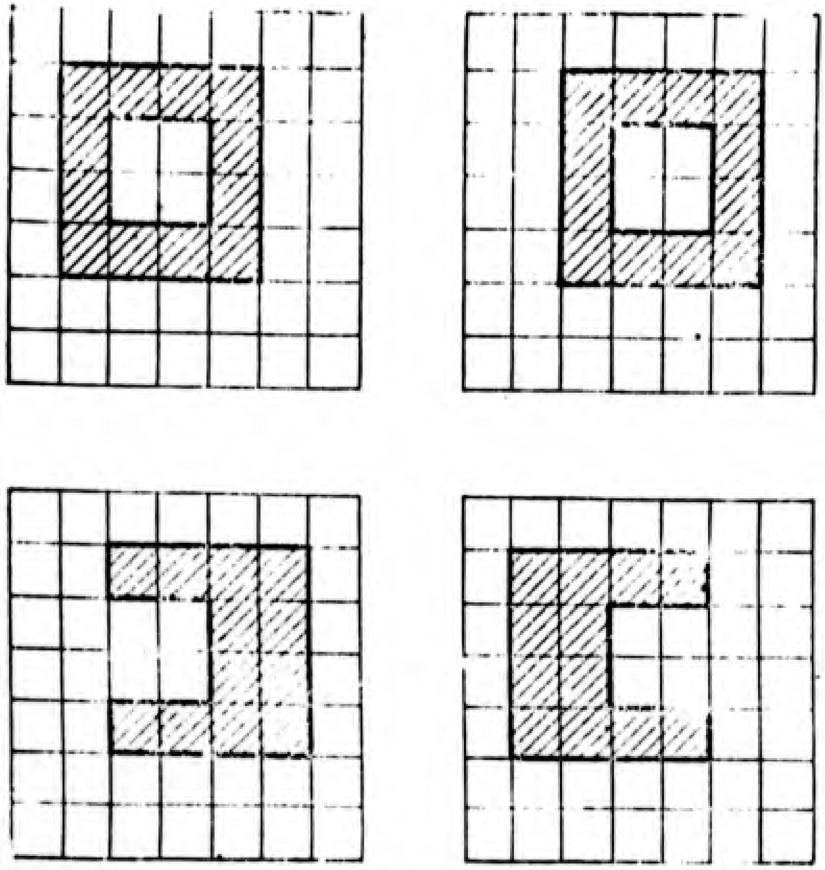


Figure 1

The point of intersection of these segments is determined from the solution to the vector equation

$$\lambda X_1 + (1-\lambda) X_2 = \delta X_3 + (1-\delta) X_4,$$

to which correspond 49 scalar equations of the following form:

$$\lambda X_{1j} + (1-\lambda) X_{2j} = \delta X_{3j} + (1-\delta) X_{4j}.$$

It is not difficult to note that in our case the indicated system has a solution when $\lambda = \delta = \frac{1}{2}$, i.e., the segments, and consequently, the convex hulls also intersect and there is not hyperplane separating set (X_1, X_2) from set (X_3, X_4).

The example considered shows that a recognition device using one hyperplane has quite limited possibilities. (This particularly pertains to the widely known "Perceptron" of F. Rosenblatt [5]). It is reasonable to assume that sets which are not separated by one hyperplane in the space of receptor signals are most probable. However, this does not mean that separation by a hyperplane is not applicable since the problem can be formulated somewhat differently. It is necessary to find a transformation of the receptor signal space in which the examined sets become linearly separable. Such a transformation is allways possible, and one of the methods of constructing it by using only oinear functions is presented by us in [6].

In this case a very serious problem arises concerning the dimensions of the space in which the examined sets become linearly separable.

From one point of view, the problem can be solved by increasing the dimensions of the space. When the dimensions of the output space are equal to the number of patterns, it can be assumed (although it is not obvious) that all patterns will be linearly independent vectors (fulfilling this condition can require construction of a space with special metrics, but we shall not concern ourselves with this problem here). If the condition of linear independence is fulfilled, we can construct a hyperplane that separates a set of patterns into two groups. It seems to us that such a solution, although possible, is unusually redundant.

From another point of view, the dimensions of the output space can be no more than $\rceil \log_2(M + K) \lceil$, where $\rceil\lceil$ denotes the smallest integer at the top, and M + K is the total number of patterns in both classes.

The latter statement seems more applicable to us. Even if, in this case, we do not manage to accomplish 100% recognition, we dom at least, ensure sufficient simplicity in the recognition unit since the dimensions of the output space never exceed 40, and is seen from the above discussion.

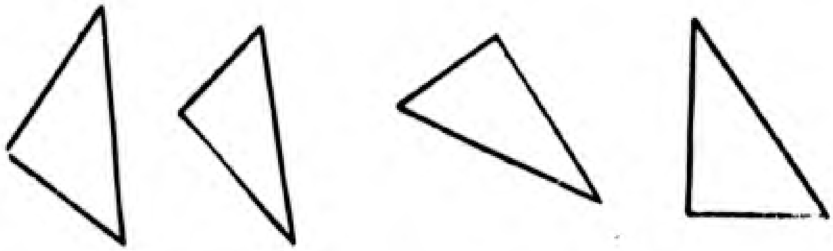
The second approach is connected with the quite complex problem of generalization (tne problem of "perception generalization" is psychology). Before going into the problem of generalization, let us consider an example.

The Paradox of Generalization

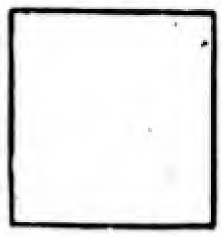
Let us assume that we are given the following two pattern sequences.

Sequence A:

Sequence B:



Then let us assume that the following pattern is given:



How should we classify this pattern in a manner consistent with the sequences given? Let us consider ourselves present during an argument between four hypothetical wise men, whom we shall call: "No-No," "Yes-No" "No-Yes," and "Yes-Yes."

"No-No": "Sequence A consists of small squares, and sequence B consists of large triangles. Since the pattern given is neigher a small equare or a large triangle, it does not belong to eigher sequence."

"Yes-No": "Sequence A consists of squares, and sequence B of triangles. Since the figure given is a square, it belongs to sequence A."

"No-Yes": "Sequence A consists of small patterns, and sequence B of large patterns. Since the figure given is large, it belongs to sequence B."

"Yes-Yes": "Sequence A consists of small squares, and sequence B of large triangles. The pattern given has properties common to both sequences, i.e., it is a square and, furthermore, it is large; therefore, it belongs to both sequences."

The wise men have presented all of the possible answers to the question as to which sequence the patterns belong, and each has offered arguments aiming to prove his own position.

Let us assume that a second group of wise men are debating as to which wise man of the first group has shown the most rational behavior.

First wise man: "The statements of all four observers concerning the sequences did not contradict the fact that they observed in sequences. They all made conclusions which were consistent with their statements; consequently, they all have shown rational behavior."

Second wise man: "Not one of the observers knew the principle of sequence formation. Consequently, they all have shown irrational behavior since they attempted to give an answer to a question without having sufficient information."

Third wise man: "In his conclusion, "Yes-No" disregarded such an important factor as dimensions. "No-Yes" disregarded such an important factor as form. Consequently, only "No-No" and "Yes-Yes", who considered all factors, have shown rational behavior."

The remaining thirteen possible philosophers of the second group continued to argue among themsleves until they had introduced all possible evaluations of the rationality of the behavior exhibited by the four eastern wise men in the first group.

The purpose of this example was not mere mental exercise, but an attempt to show the difficulties arising when we evaluate the behavior of a machine which is carrying out generalization.

A recognition learning unit must be able to generalize, i.e., to relate a selection of patterns to a set in which the given selection is contained as a subset. When generalization is accomplished in this manner, we call it sequential generalization, emphasizing by this that we do not exclude any earlier indicated pattern. In practice we can use almost sequential generalization to exclude doubtful cases which could have been erroneously classified earlier. For simplicity we shall consider only sequential generalization.

Although the single requirement we impose on generalization may be a consistency which previous experience, we are still a long way from determining a single method of generalization. Let the number of earlier demonstrated patterns be M and the number of possible patterns 2^N. As we have already noted, in all practical cases M is a negligible fraction of 2^N. The number of possible methods of generalizing a unique selection from M patterns is equal to $2^{(2^N-M)}$. On the whole, this number is many orders of magnitude more than the number of different possible patterns, which by itself is astronomically high. However, somehow, and in some cases, people rather effectively restrict the huge number of possible methods for sequential generalization to certain preferable ones.

Human generalization does not allways correspond to the simple geometric properties in a single hypercube, such as linear separability, as has already been shown above. What are the factors involved in the human's choice of reasons for generalization? The following discussion may give a hint.

The problem of generalizing a sequence of shapes is similar, in a sense, to the problem of extrapolating the function of a real variable. It is obvious that we can except sufficient probability of successful extrapolation only under the strictest limitations relative to possible extensions. Even if it is known that the unknown function is continuous, there exists an infinite set of possible extensions.

Usually the consideration of "simplicity" has a substantial effect on the choice of extension, i.e., we extend a function to the next interval so that we save ink, minimize change in derivatives, preserve smoothness, etc.

Is this intuitive supposition analogous to the case of pattern recognition? We consider it to be and shall attempt to substantiate our statement with an example. Let us assume that we are shown a sequence consisting of some small and some large squares having a variation in dimensions. Would we be right to assume that the set from which these patterns were taken includes all squares of intermediate dimensions? Of course, a set which includes squares of some size and does not include squares of other sizes is completely possible; however, the description of such a set can be quite complex. The simplest assumption, as far as description is concerned, is that all squares, regardless of size, belong to a set. We are right to make such an assumption on the same basis as a man extrapolating a segment of a straight line by the addition of another segment in the same direction. Sometimes the extrapolation is not exactly correct, but it is, at least to some extent, very simple. Let us note also that if we want to form a set consisting of squares of several different sizes, we should increase the size of the sample accordingly. If we want to construct a machine which accomplishes simple extrapolation on the basis of a short period of instruction, we must pay for this by limiting the complexity of the problems which we propose to solve.

We shall assume that in addition to this sequence requirement, generalization in pattern recognition must be characterized by simplicity. In spite of the fact that "simplicity" is a very complex concept, many agree that when describing such a concept as generalization, simplicity, to a very considerable extent, depends on the language used in the description process. In the case of patterns, the language, of course, must include geometric terms relating to shape, and methods by which different recognition units can be connected to form large structures, i.e., "vocabulary" and "grammar."

Of course, such an approach requires an evaluation of simplicity with reference to language, but as yet a generally accepted criterion for this kind of simplicity does not exist. Brevity sometimes concurs with our intuitive concept of simplicity, but this is not always true as, for example, when decimal and binary codes are compared. No less complex is the problem concerning the conditions under which it becomes necessary to expand the language by adding terms and connections and the problem of evaluating the change in simplicity occurring with this.

Thus, the introduction of the linguistic approach leads to a new problem, i.e., the problem of obtaining strict evaluations of description simplicity and comparing description simplicity in different languages.

We have already seen that the number of possible sequential generalizations with pattern selection is astronomically high; however, the above statements concerning simplicity allow us to give preference to certain possible methods over others.

This initial preference will be called "conviction."

Let us assume that a recognition learning unit does not have an inherent "conviction" in the beginning. Let us further assume that the "conviction" is formed directly from the selection. Then, during a certain period of operation with the "teacher," the recognition learning unit tends to develop a "conviction" approaching the "conviction" of its teacher." How reasonable is such an opinion?

In the first place, the number of possible "convictions" is considerable larger than the largest number we have considered so far, i.e., the number of possible generalizations. If we establish "conviction" as being of the order of generalizations, such as a simple view leads to $[2^{(2^N-M)}]!$ different "convictions."

In connection with this, there arises the problem of *a priori* information in the machine. In the case of a recognition learning unit, its structure emerges in the form of *a priori* information. Depending on initial structure, we can distinguish between highly organized recognition learning units and less organized recognition learning units. It is obvious that less organized units possess organization ability and during learning should make their structure more complex. Apparently, *a priori* information or the initial structure of a recognition unit considerably decreases the number of "convictions" from which, with instruction, selection is accomplished. *A priori* information is established by a designer on the basis of either intuitive reasoning, or strict analysis of patterns which can appear at the input of a recognition unit or analysis of the recognition process required. Factors determining *a priori* information can be intuitive, or strictly determined by the concept of "similarity" in form, or a number of other concepts which are determined when the problem of recognition is formulated and depend substantially on the character of the medium in which the machine functions or on the character of the demands made on the result of the recognition. From our point of view, we can expect success in the near fugure only when using learning machines with a sufficiently high degree of prior organization, i.e., machines possessing a large supply of *a priori* information.

Bibliography

1. Sherman, H.: A Quasi-Topological Method for the Recognition of Line Patterns, *Proceedings of the International Conference on Information Processing*, UNESCO, London, 1960.
2. Tsetlin, M. L.: Automata with Linear Tactics. *DAN*, Vol. 139, No. 4.
3. Gel'fand, I. M. and M. L. Tsetlin: The Principle of Non-local Search in Automatic Optimization Systems, *DAN*, Vol. 137, No. 2, 1961.
4. Selfridge, O. G.: Pandemonium: A Paradigm for Learning, *Symposium on the Machanisation of Thought Processes*, London, 1958.
5. Rosenblatt, F.: Two Theorems of Statistical Separability in the Preceptron, *Symposium on the Mechanisation of Thought Processes*, London, 1958.
6. Varshavskiy, V. I. and T. N. Semenova: Learning Prgoram for Pattern Recognition (printed in this collection).