ADAPTIVE PATTERN RECOGNITION OF AGRICULTURE IN WESTERN KANSAS BY USING A PREDICTIVE MODEL IN CONSTRUCTION OF SIMILARITY SETS

R. M. Haralick

Center for Research Inc., Engineering Science Division
University of Kansas
Lawrence, Kansas

ABSTRACT

Automatic classification of remote sensed data is a necessity since satellite remote sensors currently on the design board are expected to gather more data each hour than could be analyzed by traditional methods in a year. A machine which could determine the structure of the sensed environment by grouping together similar data signals and relabeling these signals with the same label would reduce both storage requirements and transmission times for the remote sensor in addition to doing pattern recognition in space or on the ground.

The paper is introduced by a discussion concerning the adaptive process of how predictive models of an environment can be generated. This discussion leads into a mathematical description of such a model for pattern recognition. The model was programmed for the GE 625 and four possible combinations of transmitted and received like and complementary polarized K-band radar images of agriculture in western Kansas were used as data.

Results confirm that various categories of vegetation such as alfalfa, grain sorghum, wheat stubble and corn cannot be distinguished from one another on the basis of their structure within the radar data. This suggests that for pattern recognition purposes, certain categories may be expanded with little or no information loss and an increase in efficiency of the pattern recognition process. Further it also suggests that if the investigator wants to distinguish between the confused categories such as alfalfa, grain sorghum, wheat stubble and corn, an additional sensor must be used which can make the distinction.

Non-imaging sensor systems such as geiger counters magnetometers, gravity gradiometers, spectrometers, photometers, and particle counters at present are collectively gathering a quarter billion bits of information per day. Some planned high-resolution imaging systems are expected to gather by themselves over three billion bits of information per day. This makes automatic pattern recognition of remote-sensed data a necessity. Over one hundred thousand analog tapes of remote sensed data are stored in ESSA's library and about two hundred additional such tapes are created daily. The only way such a great volume of information can be utilized is by automatically recognizing the pattern of relationships which exist within the data.

A pattern-recognition system is one kind of system which determines the relationship of category or type. We will endeavor to describe an adaptive pattern-recognition system, which needs no teacher, is easily built, and is able to automatically classify data. The system we describe has been simulated and tested with agricultural radar images, and the results of the test will follow the description.

A pattern-recognition system takes patterns for inputs and produces outputs which are either direct symbols for categories or binary M-tuples which form a code for categories. The input patterns form N-tuple of measurements (or sometimes their binary code) taken of some environment. Each component of the N-tuple could represent the output of one sensor system. Figure 1 illustrates the pattern recognition process.

We begin with a few elementary definitions. A system is a structured device which produces an output related to its past and present inputs according to some deterministic or probabilistic rule. The rule is called the structure of the system. If the output is related only to the present input, the system is memoryless. If the output is related to both the past and present input, the system has memory. If the form of the rule at time t depends on some of the inputs before time t, then the system is changeable; otherwise, the system is fixed.

An example of a memoryless fixed system could be a radio, the electromagnetic field at its antenna being considered the input and the acoustical field generated by its loudspeaker being considered the output. It is memoryless because the output does not depend on previous inputs: only on the present input. Past radio programs do not influence the radio's present operation. An example of a fixed system with memory could be an adding machine, the buttons pushed on the keyboard register being considered the input and the number displayed on the carriage register or paper tape being considered the output. The adding machine has memory of the sum of all the past inputs. Examples of changeable systems can be found in any of the self-learning, self-organizing or adaptive systems.

A pattern-recognition system can either have a memory or be memoryless, and can be fixed or changeable. If the system changes its own structure the change will, if it is to be meaningful, depend on the input-pattern sequence, so that the present state of structure will provide some indication of past input history; hence, a self-changeable system is a system with memory. Fixed pattern-recognition systems have limited use, since the precise transformation from input pattern to category type must be known beforehand. With the present state of the art for imaging sensors, prior knowledge is a strict requirement. Radar images, for instance, can be calibrated at best to only 3 db from the initial transmitted signal to the final developed image. Different pictures taken with the same sensor system of exactly the same environment at two different times can vary by a two-to-one gray scale variation. A fixed pattern-recognition system would have little use here, because the data has an unknown bias. All of its information is relative to the bias levels. Thus, what is needed is some kind of pattern recognition system which can change or adapt according to the biases and stochastic non-linearities of the sensor system and according to the structure of the environment sensed, so that the structure of the pattern-recognition system is optimum for the kind of job it is supposed to do.

The set of all logically possible measurements we will call measurement space. To each point in measurement space a pattern-recognition system assigns one category, so that the set of points which are assigned to each category forms a cell in a partition of measurement space, as illustrated in Figure 2.

The pattern-recognition problem is usally presented as follows: The investigator forms I training sequences S_i, $i = 1, \ldots I$ of measurements such that the measurements in the i^{th} sequence all represent measurements of objects which the investigator knows are in the i^{th} category. The i^{th} sequence actually is a sample from the i^{th} population, which is a population of objects in the i^{th} category. It is possible for the same measurements to occur in several sequences so that it is a problem to decide what category to assign to such measurements. Given the sequence S_i, $i = 1, \ldots I$, the investigator would like to find a decision rule (equivalence relation) which partitions measurement space into I cells so that the i^{th} cell contains the set of measurements which are most representative of the i^{th} category. A pattern-recognition system is a system which has such a decision rule for its structure. Since the structure of the system is determined from the training sample we may say that the system has memory relative to the training sample.

With the problem posed in this way, there is no guarantee that the categories which the investigator makes up are really representative of the structure of the measurement space. Hopefully, the categories and sensors chosen present each category as an isolated cluster in the measurement space. However, this does not always happen: sometimes there are two clusters in the measurement space which the investigator groups together as the same category. In this case, structural information of the measurement space is lost, since once categorization is done there can be no further distinctions made between these two clusters. Sometimes what the investigator calls two categories actually form only one cluster in the measurement space. In the first case the investigator is losing information, and in the second case he is wasting effort.

On the basis of these ideas we will restate the pattern-recognition problem. Given an empirical probability distribution on measurement space, we would like to find a rule (equivalence relation) which partitions measurement space so that each cell of the partition represents a cluster or similarity set in measurement space. Each point in such a set is similar to every other point in the set because they all belong to the same grouping or cluster.* Pattern recognition, approached from this perspective, becomes a method of handling data in terms of families of aggregates, each aggregate being a cluster of points.

The information which a measurement conveys to the investigator is not contained in the precise values of the components of the N-tuple, for those values, if perturbed just a little, would still usually give a measurement conveying as much information to the investigator as in the unperturbed case. Rather, the information which a measurement conveys to the investigator is contained in the aggregate family to which it belongs.

We illustrate this idea about information: suppose there are five different kinds of objects or occurrences in the environment sensed, Further suppose that to each object or occurrence there corresponds an ideal measurement, the measurement which would be made if the environment were not stochastic and if the sensor system were accurately calibrated, completely linear, and noiseless. However, due to all the conditions mentioned, measurements of the same object or occurrence will deviate a little from the ideal, distributing themselves around the ideal as illustrated in Figure 3. Since these deviations were caused by the random effects of the environment, thermal noise in the sensors, etc., the investigator is not interested in their exact values. In a classification problem the only item of interest is: which object or occurrence caused the measurement we made--i.e., which aggregate family does the measurement come from?

A pattern-recognition system which would handle the data in terms of aggregate families would have large application to the remote sensing field. Without any knowledge or training from the investigator, the system could transmit back the occurrence of an aggregate family instead of the actual measurement made. This would yield a large saving of channel capacity and preserve the information in the structure of the environment sensed.

The problem, then, is how to go about designing such a system. The key to this lies in our understanding of similarity. Similarity of measurements means that the difference between the measurements is statistically unimportant. The reason the difference is unimportant is that two measurements which are similar both belong to the same cluster (to the same aggregate family), and what characterizes the location of the cluster also characterizes the location of the measurements within it. This is the basis of similarity.

Clusters could be located by finding subsets of measurement space whose general location also characterizes the measurements within it. Location implies a given metric, however, the general problem may have a measurement space which can have no metric on it at all. The generalization of location may be easily done with measures of concomitant variation, an example of which follows. Suppose we have in mind a subset which we think is a cluster. Every time a measurement occurs which is in the subset, we will designate it "a." Every time a measurement occurs which is not in the subset, we will designate it "b." Here there is a 100% concomitant variation of the measurements in the cluster designated by the letter "a." If we wanted to remember these associations, we could store all those points designated by the letter "a" in one place in memory and those points designated by the letter "b" in another place in memory. However, if the measurement space has high dimensionality, as in multi-sensor systems, it would have so many points within it that storing all the associations would be almost impossible if not a waste of machinery and money.

Thus we divide each measurement into a number of characteristics, just as we could divide the perception of a chair into a number of characteristics which describe it. For the chair, these might be:

characteristic 1-2	-- number of legs, 3 or 4
characteristic 3	-- back or no back
characteristic 4	-- if back, is it hard or soft
characteristic 5	-- soft or hard seat
characteristic 6	-- arms or armless
characteristic 7-14	-- color of upholstery: red, orange, yellow green, blue, indigo or violet
characteristic 15-18	-- material of upholstery (cotton, leather, or plastic)

*We will use the concepts cluster and similarity set interchangeably. When the emphasis is on geometric compactness we will use the concept cluster. When the emphasis is on the similarity of the points within a cluster, we will use the concept similarity set.

Now we try to find the association between the letter "a" and each one of the characteristics. If "a" is always associated with the occurrence of the i^{th} characteristic, then a measure of concomitant variation between the letter "a" and the i^{th} characteristic would attain its maximum. If it is never associated with the occurrence of the i^{th} characteristic, then a measure of concomitant variation between the letter "a" and the i^{th} characteristic would attain its minimum. To find the total measure of concomitant variation between a measurement and the letter "a," we could appropriately sum the individual measures between each of the characteristics and the letter "a." If we wanted to compare the general characteristics of the subset (associated with the letter "a") with any one of the measurements belonging to it, we could use the measure of concomitant variation backwards We could ask: given that we have a measurement which has been assigned the letter "a," is each characteristic more likely to be associated with the letter "a" or is it more likely not be associated with the letter "a"? To answer the question, we could look at the measure of concomitant variation between the i^{th} characteristic and the letter "a." If it is larger than its neutral concomitant-variation value, then we could answer by saying that the i^{th} characteristic is indeed associated to some extent with the letter "a"; if it is less than its neutral concomitant-variation value, then we could answer by saying that the i^{th} characteristic is not associated with the letter "a." Using such a measure of concomitant variation, a system could find a cluster by finding a subset of measurement space for which the general characteristics of the subset are sufficiently related to its measurements. This relatedness is a relation of prediction or estimation.

A subset of measurement space can be called a cluster if we are able to predict or estimate the kind of measurements which are representative of it from the characteristics we have chosen. The kind of system we seek, then, is one which produces an output from which it can predict the input measurement or some similar measurement. Such a system is an adaptive system with a predictive criterion.

An adaptive system is one which examines its own output, evaluating it according to some built-in criterion, and then modifies its structure in a way which tends to improve the evaluation of the output. A predictive criterion is one which attaches high value to the ability of the system to predict or estimate the input pattern or some similar pattern, given only the output.

We now proceed to design a system of the type we just described. It is a two-layered* adaptive system with a predictive criterion, First the following definitions are in order: let $L_i = \{l_{ij}\}_{j=1}^{N_i}$, $i = 1, \ldots K$ be the set of possible quantized measurements from the i^{th} sensor in a total of K sensors. Measurement space G is the set of all possible measurements; $G = L_1 \times L_2 \times, \ldots \times L_K$. For any subset A of G, $A \subset G$, define $A_d = \begin{cases} A & \text{when } d = 1 \\ A^c & \text{when } d = -1 \end{cases}$. Construct characteristic or feature sets E_i, $i = 1, \ldots K$ such that the probability of E_i is strictly greater than zero $P(E_i) > 0$, and for every $g \epsilon G$ where $P(g) > 0$, there exists a vector $\delta(g) = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_N \end{pmatrix}$ such that $\{g\} = \bigcap_{i=1}^{N} E_{i\delta_i}$. Let the pattern set Δ be the set of all such possible vectors. A pattern here, is considered to be, not the measurement vector itself, but a binary-coded N-tuple of the measurement vector.

Of course, measurements are not naturally binary but usually analog. At some point in the recognition process, if only during transmission the analog measurement is quantized and converted to binary form. The binary code is determined by using information-theory techniques on channel capacity and noise distribution. However, the binary N-tuples which we have called patterns have little to do with such transmission codes. Rather, the binary codes which form the patterns have to do with feature or characteristic sets. We illustrate this in Figure 5, where the large rectangular area represents measurement space. The fine partition on measurement space represents the quantizing. Each measurement which occurs within a particular cell of the partition will have the same quantized value. The outlined area represents a connected subset of measurement space. Each point within the subset is close to other points within the subset. Such a subset can be called a feature or characteristic set, since it characterizes in this particular case what we can loosely call the upper-left-hand corner of measurement space. If a measurement is made which belongs to the characteristic subset, we can code the i^{th} component of the pattern corresponding to the measurement as a binary 1; otherwise, as a binary 0.

* The term layer is used in accordance with the terminology Rosenblatt developed for his perception devices. In the perception, a layer refers to a grouping of associative units which function together.

This kind of coding is quite different from a regular binary code. Suppose that the measurement vector has two components and that each component is bounded by 0 and 4 inclusive. Measurement space quantized by unit intervals is illustrated in Figure 5. A regular binary code for the quantized measurement can be formed as follows:

(0,0)	00 00	(2,0)	10 00
(0,1)	00 01	(2,1)	10 01
(0,2)	00 10	(2,2)	10 10
(0,3)	00 11	(2,3)	10 11
(1,0)	01 00	(3,0)	11 00
(1,1)	01 01	(3,1)	11 01
(1,2)	10 01	(3,2)	11 10
(1,3)	01 11	(3,3)	11 11

The first two binary bits represent the binary code for the first component while the last two binary bits represent the binary code for the last two components.

Figures 6, 7, 8, and 9 illustrate the characteristic sets which would correspond to the first, second, third and fourth binary bits in the code.

In general such codes do not have corresponding characteristics which actually characterize a connected neighborhood in measurement space. In fact, the 2nd and 4th bit just indicates whether the first and second components are even or odd. Whether a quantized measurement is even or odd does not have the slightest bearing on recognizing categories such as corn or wheat.

For any S-dimensional real vector $V = \begin{pmatrix} v_1 \\ v_2 \\ v_S \end{pmatrix}$ let the function Sgn: $R^S \rightarrow \{-1, 1\}^S$ be defined by $\text{Sgn}(V) = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_S \end{pmatrix}$ where $\mu_i = \begin{cases} +1 \text{ if } v_i - \bar{v} > 0 \\ -1 \text{ if } v_i - \bar{v} \leq 0 \end{cases}$, $i = 1, \ldots S$ and $\bar{v} = \frac{1}{S} \sum_{i=1}^{S} v_i$. Let $\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_M \end{pmatrix}$ be the binary $\{-1, 1\}$ code for the output categories. The form of the structure for the first layer of the adaptive system is defined by the equation $\eta = \text{Sgn}(Q^T \delta)$ where $Q = (q_{ij})$ is the matrix of concomitant variation coefficients and δ is the input pattern vector. The first layer is illustrated in Figure 10.

Each q_{ij} is a measure of the concomitant variation between the i^{th} pattern characteristic with the j^{th} similarity set characteristic. If it generally happens that when a pattern has the i^{th} characteristic, the similarity set into which it is classified has the j^{th} similarity set characteristic, then q_{ij} will be positive. If it generally happens that when a pattern does not have the i^{th} pattern characteristic, the similarity set into which it is classified does not have the j^{th} similarity set characteristic, then q_{ij} will also be positive. In the other two cases q_{ij} will be negative. Thus $\sum_{i=1}^{N} \delta_i q_{ij}$ is an overall measure of concomitant variation between the input pattern δ and the j^{th} similarity-set characteristic. If this sum is great enough, the system will decide that the similarity set to which it will classify δ will have the j^{th} similarity set characteristic; otherwise it will decide that it does not.

The binary $\{-1, 1\}$ vector η is the output of the system. η is a function of the input δ and the present Q matrix, which is conditioned by past inputs. The system tests its output η by forming a prediction or estimation of δ. This is done in the second layer of the system whose structural form is described by the equation $\hat{\delta} = \text{Sgn}(Q\eta)$, and is illustrated in Figure 11. The vector $\hat{\delta}$ is the system's prediction or estimation of δ, the input pattern. The output is evaluated by comparing δ and $\hat{\delta}$ component by component. The system then modifies the appropriate elements in the Q matrix in a way which tends to improve the prediction or estimation, $\hat{\delta}$. An adaptation rule of the sort $q_{ij}(t+1) = q_{ij}(t) + \epsilon\, \delta_i \eta_j \left| \delta_i - \hat{\delta}_i \right|$ can produce just such a change. Suppose $\delta_i = \hat{\delta}_i$, then there would be no change since $\left| \delta_i - \hat{\delta}_i \right| = 0$. Suppose $\delta_i = 1$ and $\hat{\delta}_i = -1$, an error situation. In order for this to happen, $\sum_{K=1}^{M} q_{ik} \eta_k$ is small; therefore, the system increases each term in the sum by an appropriate change in each q_{ik}, $k = 1, \ldots M$. To increase each term $q_{ik} \eta_k$ the system must increase q_{ik} if $\eta_k = 1$ and decrease q_{ik} if $\eta_k = -1$. Suppose $\delta_i = -1$ and $\hat{\delta}_i = 1$, also an error situation. In order for this to happen, $\sum_{k=1}^{M} q_{ik} \eta_k$ is large; therefore, the system decreases each term in the sum by an appropriate change in each q_{ik}, $k = 1, \ldots M$. To decrease each term $q_{ik} \eta_k$ the system must de- drease q_{ik} if $\eta_k = 1$ and increase q_{ik} if $\eta_k = -1$. We may summarize these changes in the table of figure 12 and compare them with the change $\epsilon \, \delta_i \, \eta_j \, | \delta_i - \hat{\delta}_i |$ produces.

An adaptive pattern-recognition system of the kind just described was simulated with the GE-625 computer, using K-band radar imagery with polarizations HH, HV, VV taken in July 1966 over agricultural areas in western Kansas. Since automatic image digitization equipment was not available, a few line scans across each field were manually taken at random with a micro-densitometer. The average of these scans was used as an estimate for the average radar return for each field. A total of 253 such fields consisting of land-usage categories of bare ground, wheat stubble, wheat stubble-weeds, wheat stubble-mulch, weeds, corn, alfalfa, grain sorghum, pasture, and sugar beets were examined.

Measurement space in this example consisted of all possible three-tuples, the first component being the average return from HH, the second component being the average return from HV, and the third component being the average return from VV. From our data the minimum, M_i, and the range, R_i, (maximum minus the minimum) for the i^{th} component, $i = 1, 2, 3$ was determined. The characteristic sets were defined as:

$$E_{ijkm} = \left\{ X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \middle| \; M_i + \frac{jR_i}{10} \leq x_i \leq M_i + \frac{(j+4)R_i}{10} \text{ and } M_k + \frac{mR_k}{10} \leq x_k \leq M_k + \frac{(m+4)R_k}{10} \right\}$$

$i = 1, 2$; $j = 0, \ldots, 5$; $k = i + 1, 3$; $m = 0, \ldots, 5$. The geometric configuration of the characteristic sets are displayed in figure 13.

The adaptive pattern-recognition system worked as follows. For each iteration a measurement from the set of 253 measurements was chosen at random. Given the definition of the characteristic sets, the system determined the δ vector which corresponded to the randomly chosen measurement. Via the first layer of the adaptive system, η, a 3 x 1 binary $\{-1, 1\}$ vector, was computed from Sgn $Q'\delta$. Via the second layer of the adaptive system $\hat{\delta}$, the prediction of δ given η, was computed from Sgn $Q\eta$. The system then compared δ and $\hat{\delta}$ and reinforced the elements q_{ij} of Q appropriately and started another iteration.

The reinforcement parameter ϵ was chosen at .002. Values of .009 and .02 were also tried but they made little difference in the final classification. (There was a random shift of at most seven points with the different values of ϵ.) The cycling of iteration after iteration was kept up until the total classification for all the data points, based on the updated Q matrix, did not change for more than 30 cycles. It took 189 cycles to reach this situation.

Figure 14 illustrates a scattergram of the data, and figure 15 illustrates a scattergram of the data coded by their respective land-usage categories as determined from ground truth. The axes of the scattergrams are the first two normalized eigenvectors of the covariance matrix for the data. These axes are usually called the first two principal axes. A frequency chart, as determined after 189 cycles, of similarity sets versus the land-usage categories is shown in figure 16. Similarity sets II and III are closely related, their difference mainly being that III has all the corn while II has most of the wheat stubble-weeds. For illustration and basic interpretation purposes we group together similarity set II and III as in the chart of figure 17. If we now group together those categories which appear to occur in mostly the same similarity sets, we will obtain the chart in figure 18. Scattergram of the similarity sets (clusters) which the adaptive system determined is shown in figure 19.

Bare ground had formed a cluster of low returns; wheat stubble, wheat stubble-weeds, wheat stubble-mulch, and pasture formed a cluster of medium-low returns which meshed with each other almost completely and overlapped with bare ground; alfalfa, grain sorghum, weeds and corn formed a cluster of medium returns; and sugar beets formed a cluster of high returns. The chart in figure 20 shows the conditional probability of the category group given the similarity set. The purpose of this chart is to arrive at some quick intuitive interpretation of the goodness of the results. Basically the chart indicates that the similarity sets correspond to the category groups with about an 85% probability.

These results show that, without a priori knowldege, categories of vegetation such as:

(1) wheat stubble, wheat stubble-weeds, wheat stubble-mulch, and bare ground
(2) grain sorghum, corn, weeds, and alfalfa

cannot be distinguished from one another solely on the basis of their structure within the K-band radar data taken during the month of July. However, from empirical knowledge, we know the returns from bare ground are probably less than the returns from wheat stubble, wheat stubble-weeds and wheat stubble-mulch; therefore, we may further partition the cluster of low returns into two parts, the first part being bare ground and the second part being the wheat stubble, etc. Also, we know that the returns from alfalfa are probably less than the returns from corn, so that we may partition the cluster of medium returns into two parts, the first part being alfalfa and the second part being corn. It is impossible to separate grain sorghum and weeds from any part within their cluster, so if they must be recognized, an additional sensor must be used which can detect them separately from the returns of other categories.

The confounding of categories within a cluster leads to the conclusion that a pattern-recognition system of the type we have described should preserve to some extent the information contained by the locality of the pattern within the cluster. This leads to a pattern-recognition system which finds clusters and then sets up a principal coordinate axis for each cluster. The output of such a system would consist of two pieces of information, the first being the cluster in which a pattern occurs and the second being the projection of the pattern onto the principal axis of that cluster. Figure 21 illustrates the principal axis idea. It would allow for a natural categorization by the clusters and a finer categorization using the principal axes, where the finer categorization depends on our empirical knowledge of the environment sensed.

The operation of such a system would closely correspond to the way our perception system works. When we see an environment we first see the environment in an appropriate general frame of reference. Then we examine the environment in fine detail from the perspective of this frame of reference. The appropriate general frame of reference is analogous to the determination of cluster (similarity set). The finer examination is analogous to the determination of the location of pattern relative to the cluster (similarity set), i.e. the projection of the pattern on the principal axis for the cluster.

ACKNOWLEDGMENTS

Financial support for this study was provided by Project Themis (DoD), DAAK02-68-C-0089; the financial support and cooperation of the University of Kansas Computation Center is also gratefully acknowledged.

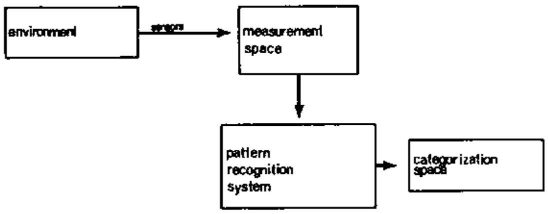

FIGURE 1. PATTERN RECOGNITION PROCESS

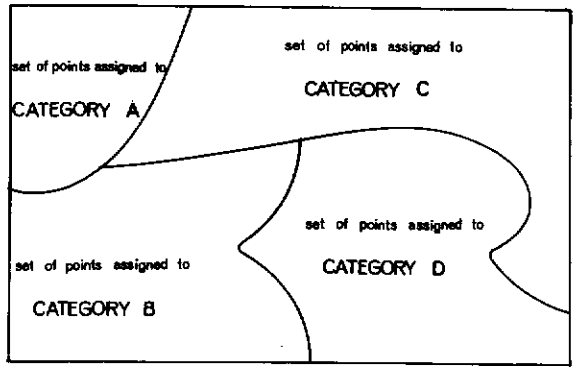

FIGURE 2. MEASUREMENT SPACE AND CATEGORIES

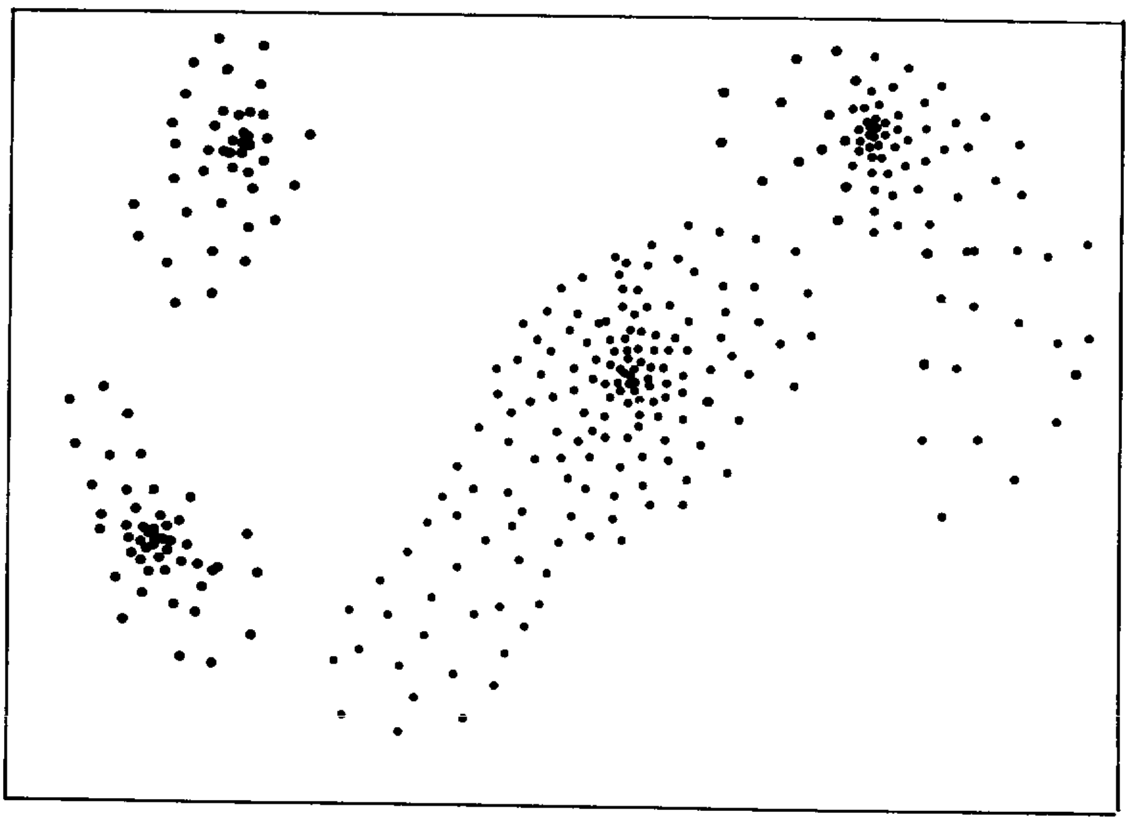

FIGURE 3. SCATTERGRAM OF MEASUREMENTS OF STOCHASTIC ENVIRONMENT WITH FOUR OBJECTS

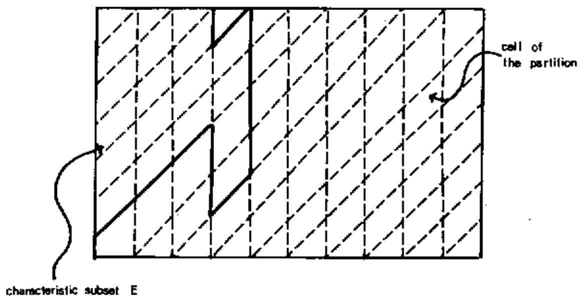

FIGURE 4. MEASUREMENT SPACE QUANTIZED

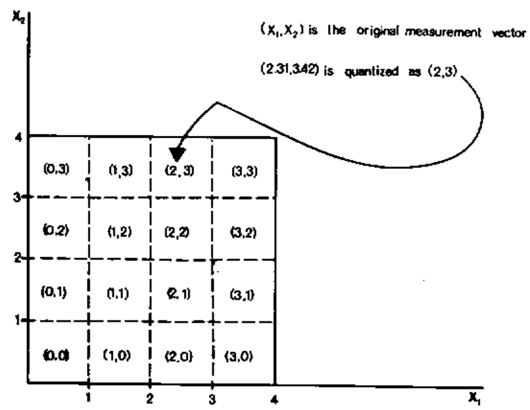

FIGURE 5. MEASUREMENT SPACE QUANTIZED BY UNIT INTERVALS

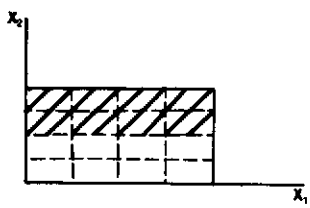

FIGURE 6. CHARACTERISTIC SET FOR FIRST BINARY BIT

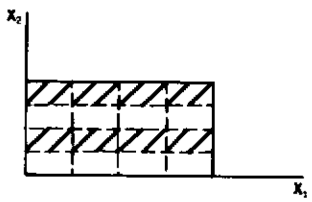

FIGURE 7. CHARACTERISTIC SET FOR SECOND BINARY BIT

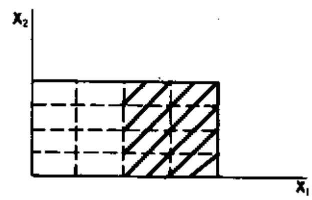

FIGURE 8. CHARACTERISTIC SET FOR THIRD BINARY BIT

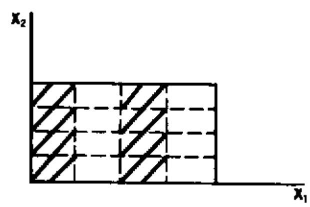

FIGURE 9. CHARACTERISTIC SET FOR FOURTH BINARY BIT

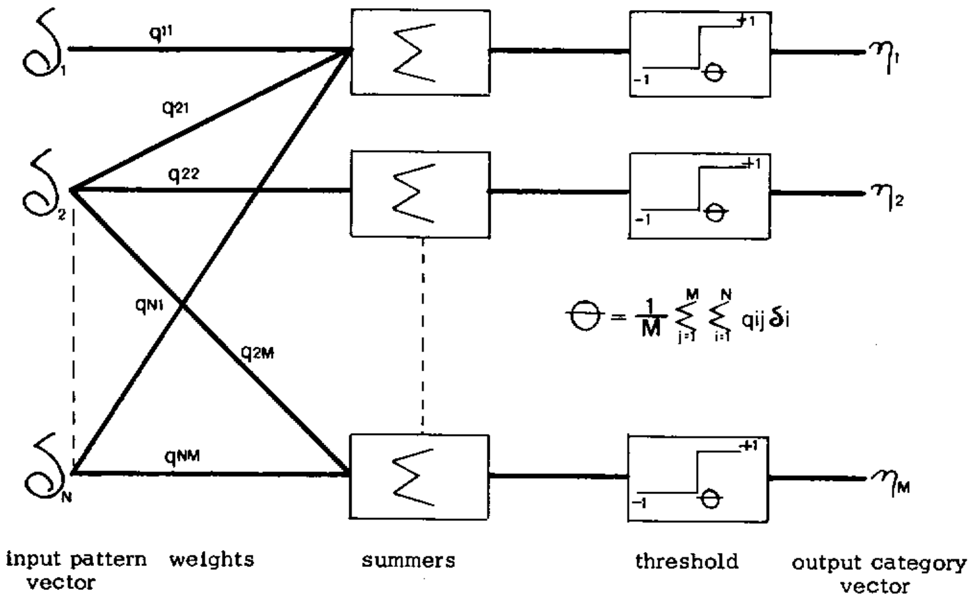

Figure 10. First Layer of Adaptive System

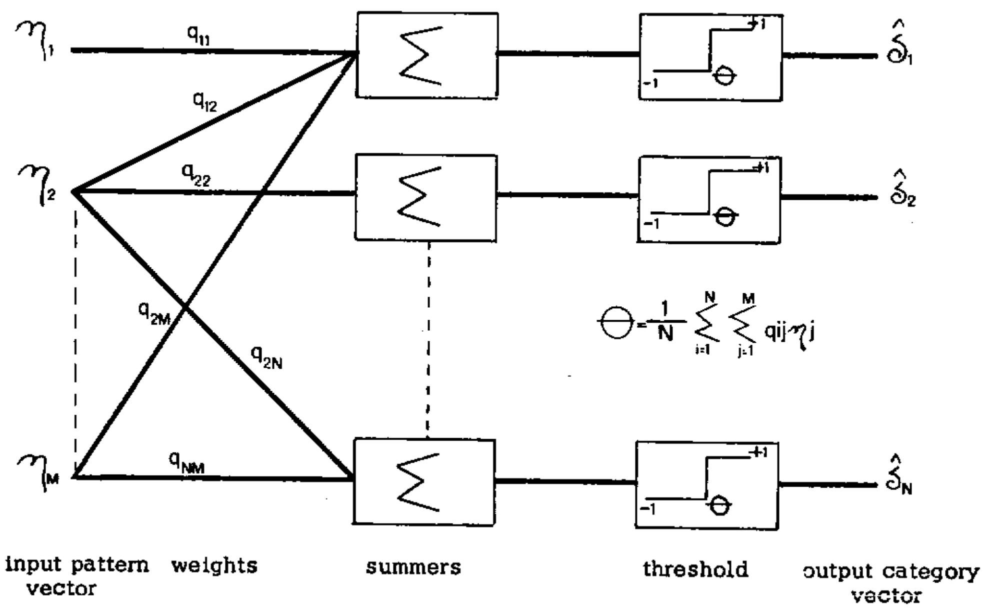

Figure 11. Second Layer of Adaptive System

δ_i	δ_i	η_j	CHANGE REQUIRED FOR q_{ij}	CHANGE PRODUCED BY $\epsilon \delta_i \eta_j \lvert \delta_i - \zeta_i \rvert$
1	1	1	none	0
1	1	-1	none	0
1	-1	1	increase	2ϵ
1	-1	-1	decrease	-2ϵ
-1	1	1	decrease	-2ϵ
-1	1	-1	increase	2ϵ
-1	-1	1	none	0
-1	-1	-1	none	0

Figure 12. Comparison of Change Required By and Change Produced By Reinforcement Statement

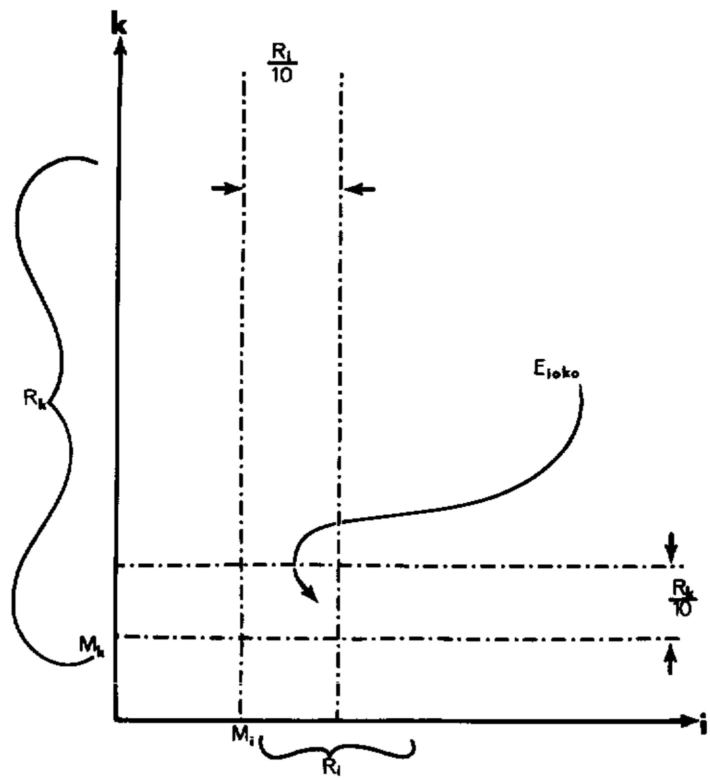

Figure 13. Geometric Configuration of a Typical Characteristic Set

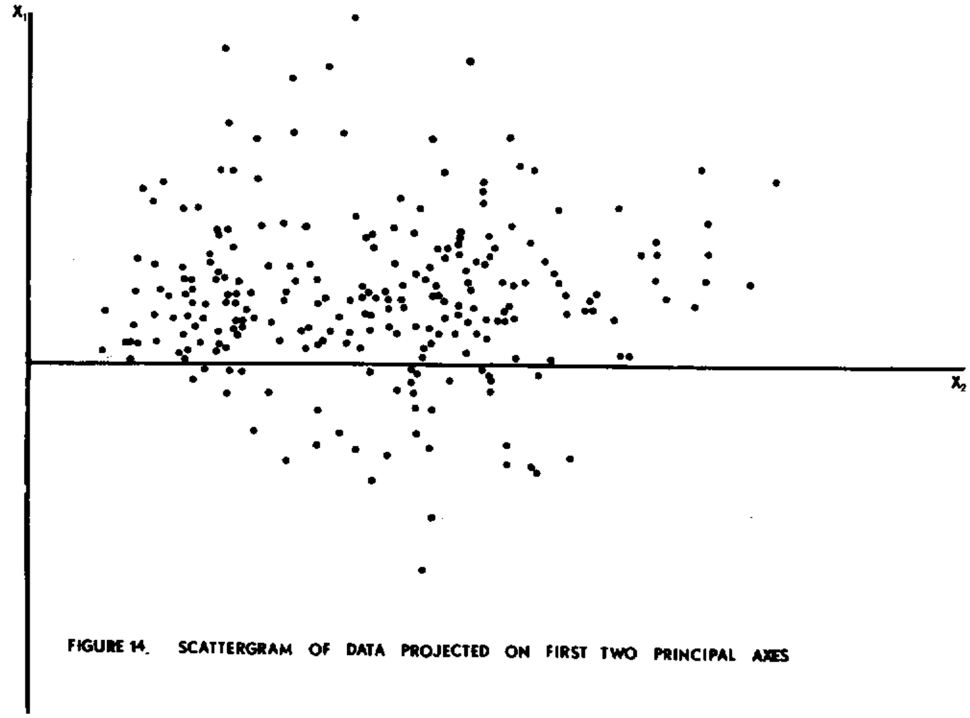

FIGURE 14. SCATTERGRAM OF DATA PROJECTED ON FIRST TWO PRINCIPAL AXES

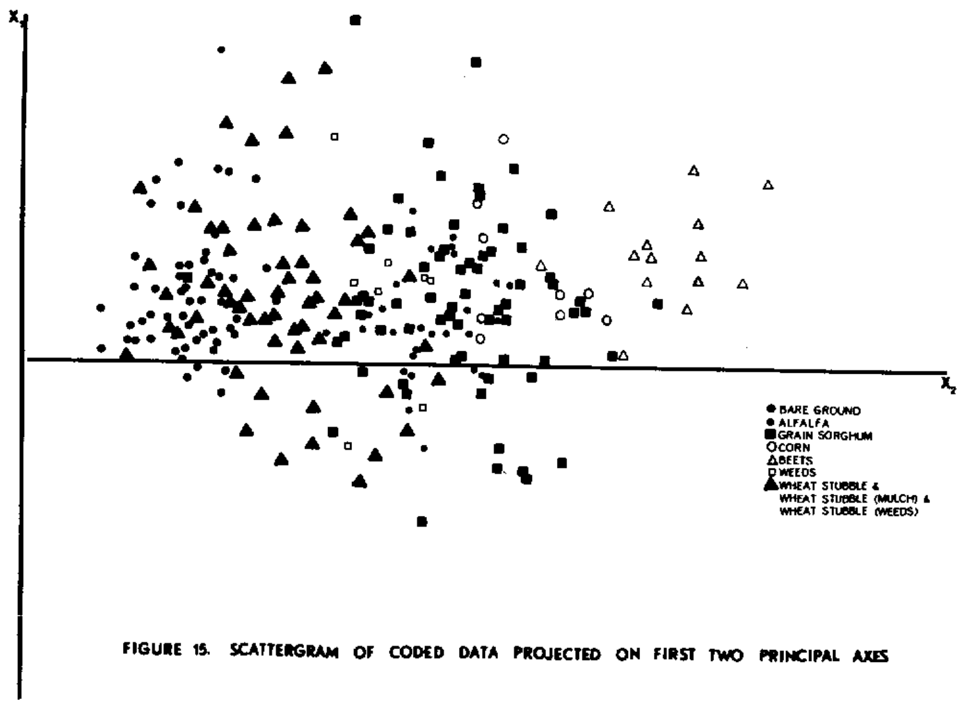

FIGURE 15. SCATTERGRAM OF CODED DATA PROJECTED ON FIRST TWO PRINCIPAL AXES

CATEGORY	I	II	III	IV
	SIMILARITY SET			
BARE GROUND	52	3	0	0
CORN	0	0	7	2
WHEAT STUBBLE WEEDS	10	22	2	0
ALFALFA	0	14	15	0
GRAIN SORGHUM	1	22	47	1
WHEAT STUBBLE MULCH	9	3	0	0
WHEAT STUBBLE	9	5	0	0
WEEDS	0	5	4	0
PASTURE	1	3	2	0
SUGAR BEETS	0	0	3	11

FIGURE 16: CROSS TABULATION OF CATEGORY SIMILARITY SET FREQUENCIES

CATEGORY	I	II + III	IV	
	SIMILARITY SET			
BARE GROUND	52	3	0	
CORN	0	7	2	
WHEAT STUBBLE WEEDS	10	24	0	
ALFALFA	0	29	0	
GRAIN SORGHUM	1	69	1	
WHEAT STUBBLE MULCH	9	3	0	
WHEAT STUBBLE	9	5	0	
WEEDS	0	9	0	
PASTURE	1	5	0	
SUGAR BEETS	0	3	11	

FIGURE 17: CROSS TABULATION OF CATEGORY AND SIMILARITY SET FREQUENCIES

CATEGORY	I	II + III	IV	
	SIMILARITY SET			
GROUP A BARE GROUND WHEAT STUBBLE (MULCH) WHEAT STUBBLE	70	11	0	
GROUP B CORN WS WEEDS ALFALFA GRAIN SORGHUM WEEDS PASTURE	12	143	3	
GROUP C SUGAR BEETS	0	3	11	

FIGURE 18: CROSS TABULATION OF CATEGORY GROUP SIMILARITY SET FREQUENCIES

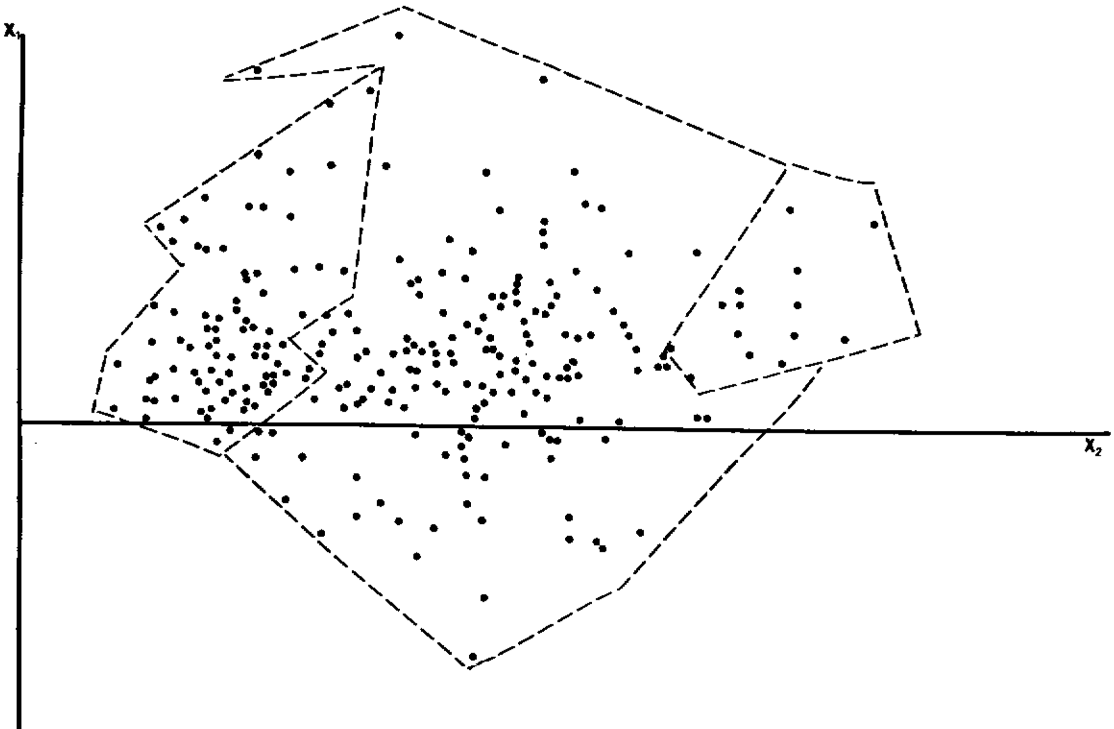

FIGURE 19. CLUSTERS THE ADAPTIVE PATTERN RECOGNITION SYSTEM FOUND

CATEGORY	SIMILARITY SET I	II + III	IV	
GROUP A BARE GROUND WHEAT STUBBLE MULCH WHEAT STUBBLE	.854	.070	0	
GROUP B CORN WS WEEDS ALFALFA GRAIN SORGHUM WEEDS PASTURE	.146	.910	.214	
GROUP C SUGAR BEETS	0	.02	.786	

FIGURE 20: CONDITIONAL PROBABILITY OF CATEGORY GROUP GIVEN SIMILARITY SET

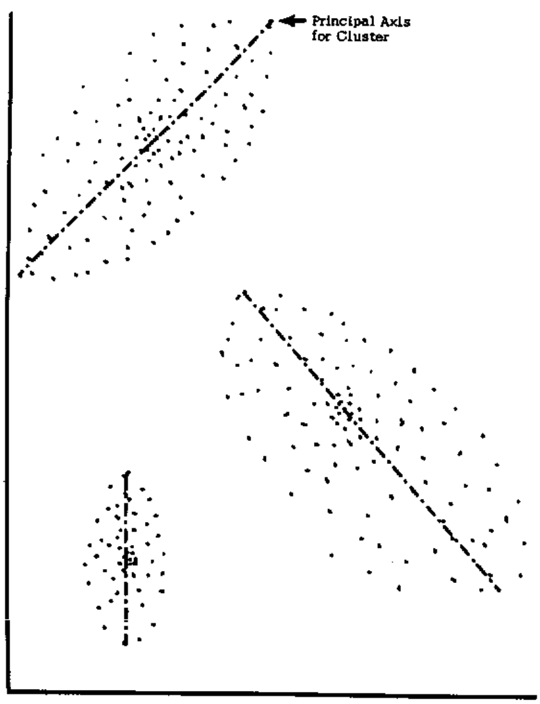

Figure 21. Clusters and Principal Axis Idea