General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

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

CRES TECHNICAL REPORT 118-5

By

Robert Haralick
Remote Sensing Laboratory
University of Kansas
Lawrence, Kansas

October 1967

Prepared by the University of Kansas Center for Research, Inc., Engineering Science Division, Lawrence, Kansas for the National Aeronautics and Space Administration (NASA) under NASA Contract No. NSR 17-004-003 and NAS 9-7175.

MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

Technical Report
118-5

ADAPTIVE PATTERN RECOGNITION BY USING A PREDICTIVE MODEL IN CONSTRUCTION OF SIMILARITY SETS*

by

Robert Haralick

October 1967

The Remote Sensing Laboratory

*The Research reported on here has been sponsored by the National Aeronautics and Space Administration under contract NSR 17-004-003 and NAS 9-7175. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Adaptive Pattern Recognition By Using a Predictive Model in Construction of Similarity Sets

Abstract

Automatic classification of remote sensed data is a necessity since satellite remote sensors currently on the design 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 of remote sensor in addition to doing pattern recognition in space or on the ground.

An adaptive predictive model is a model which generalizes and induces the similarity set (category) which a sample of given data belong and then assuming given the similarity set from which the data came it predicts or deduces the characteristics of the data. When the predictions are in error the parameters in the model adapt to make it more likely that the predictions will be correct for the next sample of data. By continuing this process iteratively, it should be possible to generate the natural categories or similarity sets which reflect the structure of the environment.

The paper is introduced by a discussion concerning how adaptive predictive models of an environment can be generated. This discussion then leads into a precise mathematical description of such a model using as the similarity criterion the ratio of the conditional odds ratio.

One of the most striking aspects of life is its organization. Life can only exist in an ordered fashion in a world where there are continually repeated and relatively unchanging sequences of phenomena (i.e. stationary and/or ergodic processes). The general pattern recognition problem involves finding what are the continually repeated and relatively unchanging sequences in measurements of an environment, whether this environment be sets of photographs, voice recordings, scientific data or the stock market.

Man has been consciously or unconsciously concerned with pattern recognition from time immemorial. Primitive man was concerned with agriculture: how to get enough to eat. This meant recognizing and using successful agricultural patterns. Biblical man was concerned with ethics: how to act in a way best to increase the harmony of the community. The patterns of behaving which did increase the harmony and organization of the community came to be known as the commandments. The early Greeks were concerned with understanding the patterns of the world: how to find the regularities and organization of the world. The things which Plato saw as steady and regular he called "ideas." Today we are still concerned with understanding the world, but our understanding has become so complicated that we need automated machines which can develop, so to speak, their own understanding of the world. The categories or concepts in which an understanding of the world is embodied are so complex, and the scientific measurements from which the characteristic or typical patterns must be found, are so numerous that man is not adequately equipped to undertake the pattern recognition process alone. Thus in the last few years a search has been undertaken to describe abstractly the pattern recognition process in the hopes that such a description will aid both man and machine to better interpret and understand the world.

What is the pattern recognition process? Suppose we are given some environment, a particular part of the real world, in which we are interested. Our interest might be in finding specific geological rock formations or faults, finding expected crop yield, determining behavioral models of groups of people or societies, or recognizing camouflaged objects such as tanks and guns. Depending upon our interests, we make appropriate measurements of the

environment, i.e. we record patterns (a pattern is any ordered set of items, facts, or measurements from a pattern group), and we examine the resulting pattern samples, classifying similar patterns as the same and/or recognizing to which pre-assigned category each sampled pattern is most similar. This examination of patterns we call the pattern-recognition process. Usually the environment is complex, allowing a multitude of possible non-dependent patterns. When the investigator says he would like to know the relationships which exist between the patterns--which patterns generally occur together, which patterns follow one another, etc. he is, in effect, saying that he is interested in finding the regularities which reflect themselves in the organization of the environment. When the investigator says he would like to understand the structure of an environment, he is searching for the respective kinds of regularities in the environment, i.e. similarity sets.

Similarity sets are the equivalence classes formed by a similarity equivalence relation in the pattern group. Thus the similarity sets form a partition over the pattern group. Each pattern belonging to a similarity set is similar to or means the same as any other pattern in the similarity set. The fundamental problem is, then, to define precisely what this similarity equivalence relation is. Since different investigators call different kinds of relations similar, we are at a loss to choose a unique definition. However, it also might be that different investigators would respond to a common denominator of similarity in a given pattern group, even though they themselves remain unaware of it. If this common denominator exists, then we may be able to define a similarity equivalence relation so that its equivalence classes form a refinement of the partitions formed by other similarity equivalence relations. Thus we would have similarity sets which are precisely defined and independent of the particular interests of the investigator.

What is similarity? We can talk about similarity in at least two useful ways: perceptual similarity or equivalence similarity. By perceptual similarity, we mean here similarity with respect to various transformations, such as rotation, translation and scaling. By equivalence similarity, we mean similarity with respect to prediction. Given that a pattern P occurs, what other

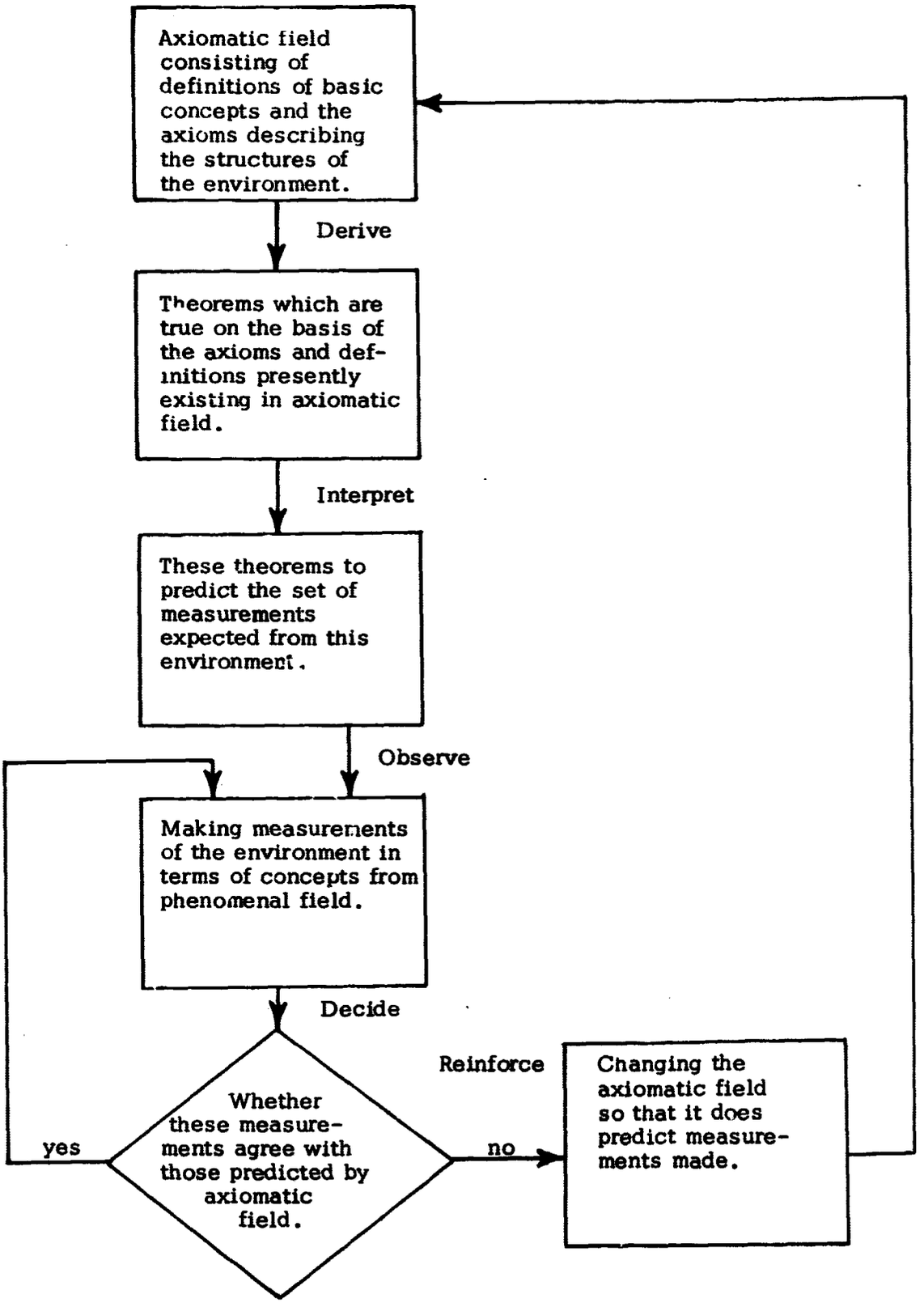


Figure 3: Flow Chart for Adaptively Constructing Axiomatic Field

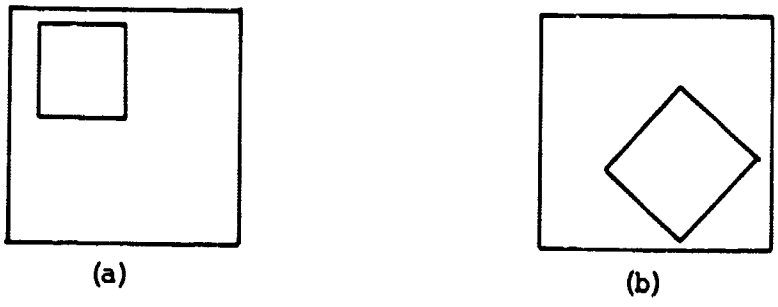


Figure 1.

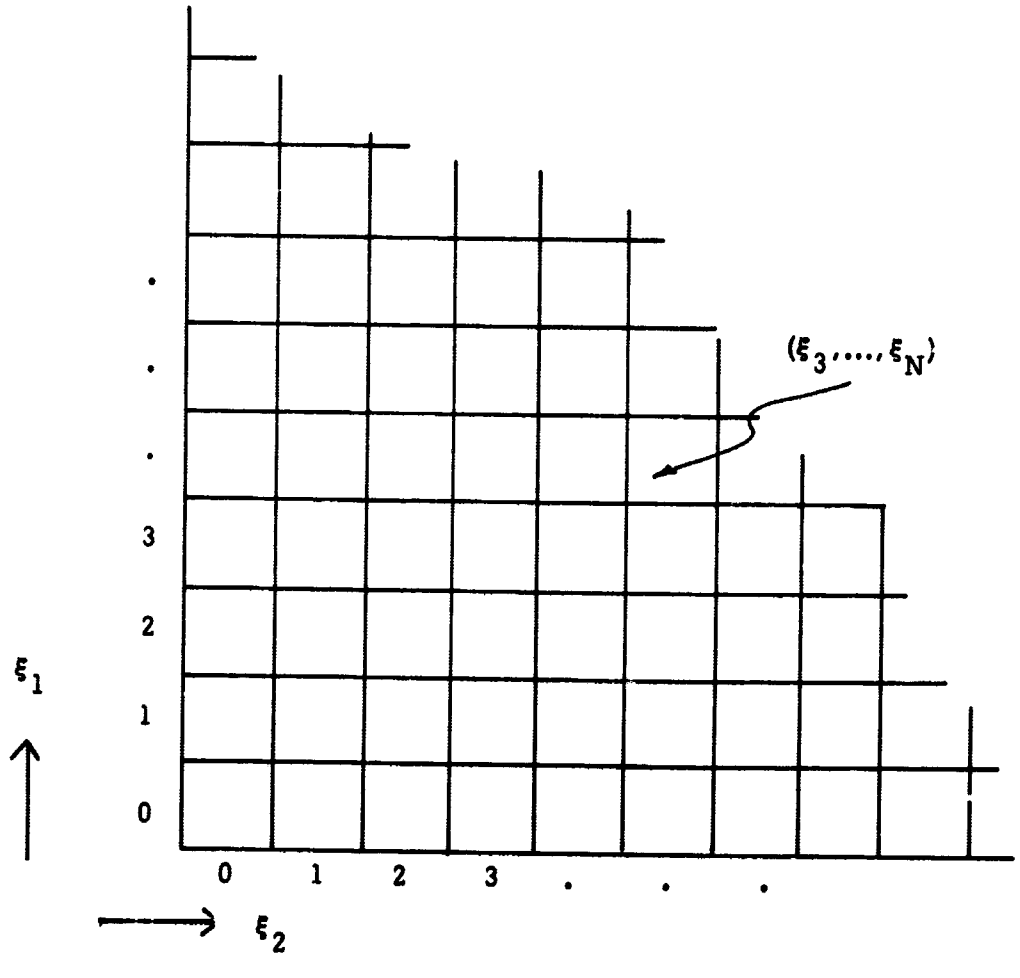


Figure 2.

the relative frequency of the patterns determined from a sample of S. Or suppose we are observing the behavior of an organism. We first determine the kind of actions in which we are interested, watch the organism and after every interval of time-1 sec, 1 minute, etc.-we record its actions. After M intervals of time we have recorded a sequence $S = \{s_i\}_{i=1}^{M}$ where $s_i = (\xi_1 \ldots \xi_N)$ and each ξ_i is a measurement of the i^{th} action in which we are interested. In either case, we would like to formulate a concept which can tell us which patterns $(\xi_1, \xi_2, \ldots \xi_N)$ are similar by equivalence. To do this we must now rephrase the pattern recognition problem to see its relationship to intelligent behavior and to abstracting. We will then see in what way they are all founded on similarity.

So far the best objective definition of intelligent behavior has been the behavioral one.[1] Briefly it goes like this: if an organism acts in an effective way to achieve a relatively complicated goal, we call its behavior intelligent. Thus, if a six-year-old child can solve elementary algebra problems, we say the child is intelligent. In ordinary everyday usage, intelligent behavior implies self-actuated, internal goal-directed, self-organizing behavior. The child who can work algebra problems has not gotten this ability out of thin air. He has been trained or educated; he has been directed by his teacher or his environment. Either he had been shown how to work the problems or his environment provided him with enough clues so he could figure out how to work the problems. In either case, the fact that he could work algebra problems is surprising and unexpected. It is this unexpected ability to perform a relatively complicated task which we usually call intelligent behavior. Thus intelligent behavior is relative to average behavior. We call behavior intelligent when it is more effective than the currently existing average behavior which cannot solve relatively complicated tasks.

---

[1] Hartley and Hartley, Outside Readings in Psychology, Thomas Crowell Company (New York, 1957), p. 243.

Now why is it that complicated tasks exist, which have a solution, but which can not be solved now? By what process do we eventually solve problems which are presently too complicated to be solved? Clearly, the answers to both these questions are closely related.

Let us take some examples. Suppose our environment is the surface of the earth, with all the processes which exist upon it. Our interest is transportation; we would like to go by car from A to B but we cannot, since we do not know which roads to take--nor, in fact, where the roads are which we could take. We solve the problem by obtaining, by means of our own research or by others' experience, a map which is a homomorphic copy of the earth's surface, and, with respect to main roads and highways of interest, an isomorphic copy of the earth's surface. Once we have such a map our route is a simple problem

Suppose our environment is mathematics and we would like to learn about algebra. Again we wish to obtain a map. This time the map is a set of definitions and axioms by which the theorems of algebra can be derived. A map in the form of definitions and axioms is a compact isomorphic copy of the world of algebra. In this case the map may not be unique, since there may be equivalent axioms and/or definitions which lead to the same theorems.

Suppose our environment is people, and we are interested in group behavior. Again we seek a map. Here the map is usually some sort of model describing the group, and is thus an isomorphic copy of it.

Suppose our environment is the universe. We send rockets with various remote sensors to sense the planets, stars, galaxies etc. The rocket telemeters back huge amounts of data that we wish to analyze. To analyze the data means to reduce it, determining the different objects which it represents. Again, these objects, recognized and placed according to their coordinates, map the universe and are a homomorphic copy of it. This problem of analysis is of course the problem of pattern recognition.

In each case, in order to achieve a solution to the problem, we construct a map which is homomorphic to the environment and isomorphic to the process of interest. Once we have such a map, straightforward applied engineering can solve the problem. Now we are in a position to answer our two initial questions. Relatively complicated problems exist precisely because we do

not have at hand an isomorphic copy of the kind of processes (if such processes exist) by which the problem may be solved. Thus, we solve such problems by obtaining an isomorphic copy of the processes of interest.

The big problem, of course, is to discover how we can obtain or construct a map homomorphic to the environment and isomorphic to those processes and things in the environment of interest. We can do this by abstracting.

We generally begin abstracting from a field of imprecise, elementary, irreducible characteristics, used daily but usually not well understood. This field of characteristics, which we call the phenomenal field, enables us to describe our environment as we perceive it or as we measure it. Examples of some of these characteristics might be velocity, temperature, charge, mass, shape, permittivity, permeability, reflection of electromagnetic waves as a function of frequency, radioactivity, material strength, chemical composition, etc. The process of abstracting develops from our phenomenal field an axiomatic field, which is a field of irreducible characteristics, definitions of concepts involving these characteristics, and axioms which show the assumed relationship between the characteristics. The basic property of the axiomatic field is that from it we can derive a set of theorems which are a description (a homomorphic copy) of the environment.

The set of axioms, definitions, and derived theorems constitutes what we usually call an explanation of the environment. Such an explanation is no more than the coherence or agreement of our observations, which are made by measuring the elementary irreducible characteristics, with the predictions made by the derived theorems. In other words, we explain our environment when we have a set of principles which are descriptions of the order and structure (i.e. constraints) in the environment.

In the process of abstracting, we adaptively develop, on the basis of past and current observations and our present axiomatic field, a new axiomatic field which better explains the environment. The reason for such a development is that we find our present axiomatic field inconsistent or incomplete: that is, the copy of the environment derived from the axiomatic field may be logically contradictory or homomorphic when we wish it to be isomorphic with respect to our interests. We illustrate this in Figure 3, where a new axiomatic field is developed by:

1) deriving theorems which are true on the basis of the axioms presently existing in the axiomatic field;

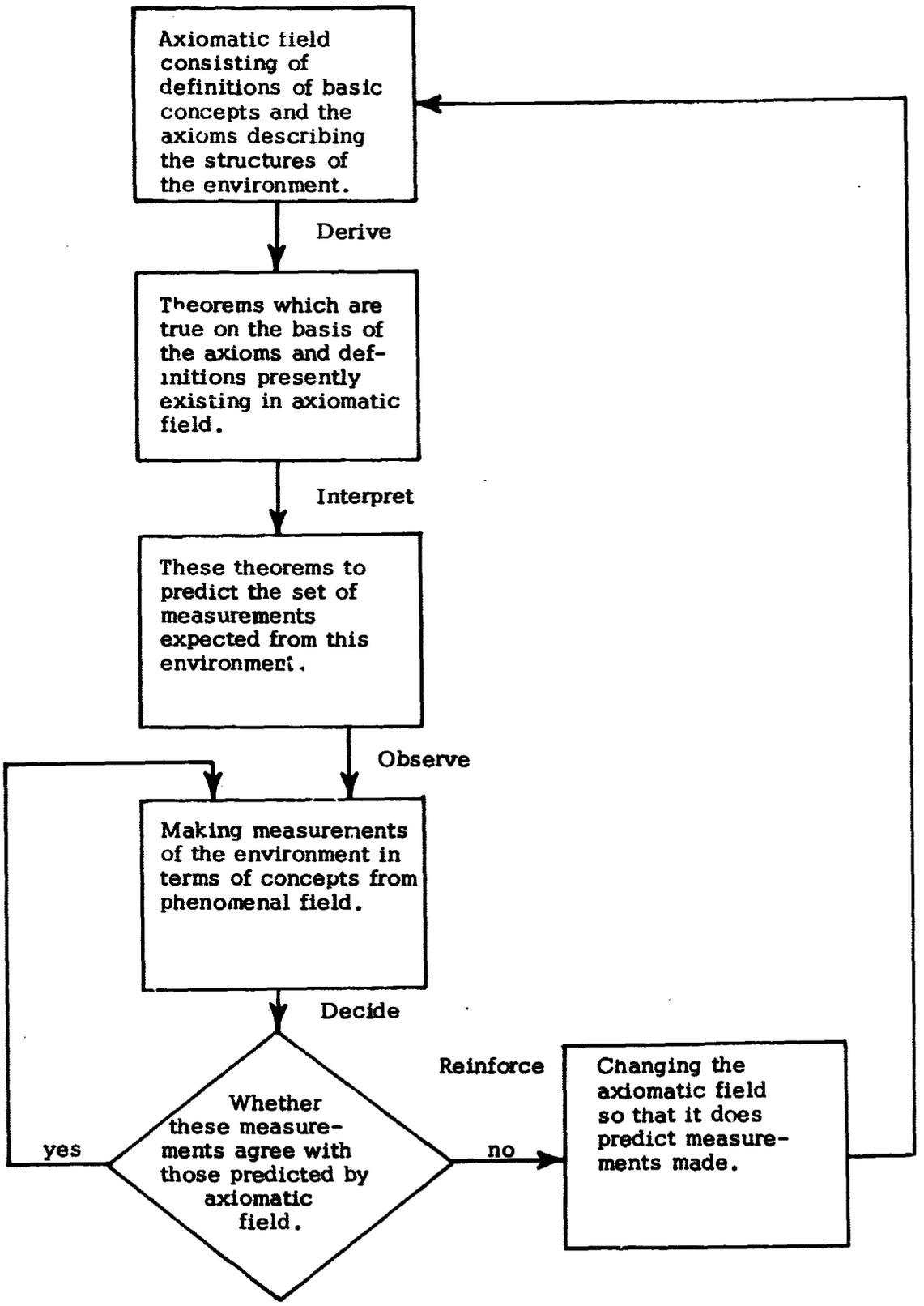


Figure 3: Flow Chart for Adaptively Constructing Axiomatic Field

2) interpreting these theorems to predict other measurements of the environment, to which the axiomatic field is to be isomorphic;

3) observing the environment, to make new measurements; and

4) changing the axiomatic field, if necessary, so that all the acceptable measurements made are measurements which have been predicted.

A survey into the history of science easily reveals this kind of process. We examine one instance: consider the phlogiston theory widely held in the eighteenth century to be the axiomatic field.

> The phlogiston theory..gave order to a large number of physical and chemical phenomea. It explained why bodies burned--they were rich in phlogiston--and why metals had so many more properties in common than did their ores. The metals were all compounded from different elementary earths combined with phlogiston, and the latter, common to all metals, produced common properties. In addition, the phlogiston theory accounted for a number of reactions in which acids were formed by the combustion of substances like carbon and sulfur. Also, it explained the decrease of volume when combustion occurs in a confined volume of air--the phlogiston released by combustion spoils the elasticity of a steel spring.[2]

For a great number of instances, the phlogiston theory was an adequate description or map of actual reality. However, when men from Black through Scheele claimed the phlogiston theory to be an adequate map for all actual reality, anomalies appeared; for example, some objects increased in weight when burned. Thus, when the phlogiston theory was interpreted to predict certain measurements and observations which did not in fact occur, scientists were led to do further research and thinking which eventually led to the oxygen theory of combustion, a new axiomatic field.

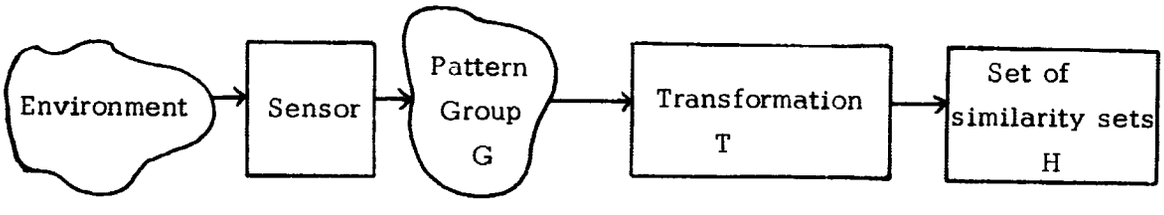
To summarize: the key in understanding the abstracting process lies in our understanding of the structure of the axiomatic field. The structure of the axiomatic field is homomorphic to the structure of the environment. In a conceptual sense, we have a partial copy of the environment in the axiomatic field. The process of abstracting is an adaptive translating process from the set of measurements to a formal axiomatic language which describes these measurements. The important characteristic of this translating process is that it preserves similarity. Similarity is preserved when the axiomatic

---

[2] Kuhn, Thomas S., The Structure of Scientific Revolutions, The University of Chicago Press, (Chicago, 1962), p. 98.

field leads to theorems which indeed predict the set of currently available measurements from the environment.

We now more precisely formulate a definition of the pattern recognition process.



Model of Pattern Recognition Process

An environment is measured with various sensors. The set of these measurements we define as the pattern group G. The pattern recognition process is a transformation T of G onto H, $T:G \rightarrow H$, where H is a set of similarity subsets of G. The order of the set G is much larger than the order of the set H, and, for its order, H represents the best copy of the set of patterns G. Transformation T is a decision rule which minimizes the expected risk.

Some examples of this process would be the Baye s' Decision approach and the Linear Hyperplane approach. In either case the investigator beforehand sets up training regions or finds good patterns typical of the categories he is interested in. This immediately limits the effectiveness of the pattern recognition process, since it is not known beforehand what the set H should consist of if it is to be a representative copy of G. Since the transformation T in the Linear Hyperplane case is a linear transformation, even if good representative patterns have been chosen, linear hyperplanes will not be the best boundaries between different similarity sets--unless G is such that the best copy of it consists of a set of convex subsets. In the Baye's Decision rule case, further a priori assumptions on the probability distribution of the similarity sets are required. It would be nice to have both a minimum-risk decision rule, such as the Baye's approach, with the ability to adaptively construct the set H so that it is a good copy of G.

We will now try to apply our insight of adaptive abstracting to the pattern recognition process. On the basis of the flow chart for the adaptive abstracting process (Figure 3), we might suggest the flow chart shown in Figure 4 for adaptive pattern recognition.

One of the problems with a probability matrix, as shown in Figure 4, is its storage problem. Storage requirements grow exponentially as the dimensionaligy of the pattern increases linearly. This can be seen immed-

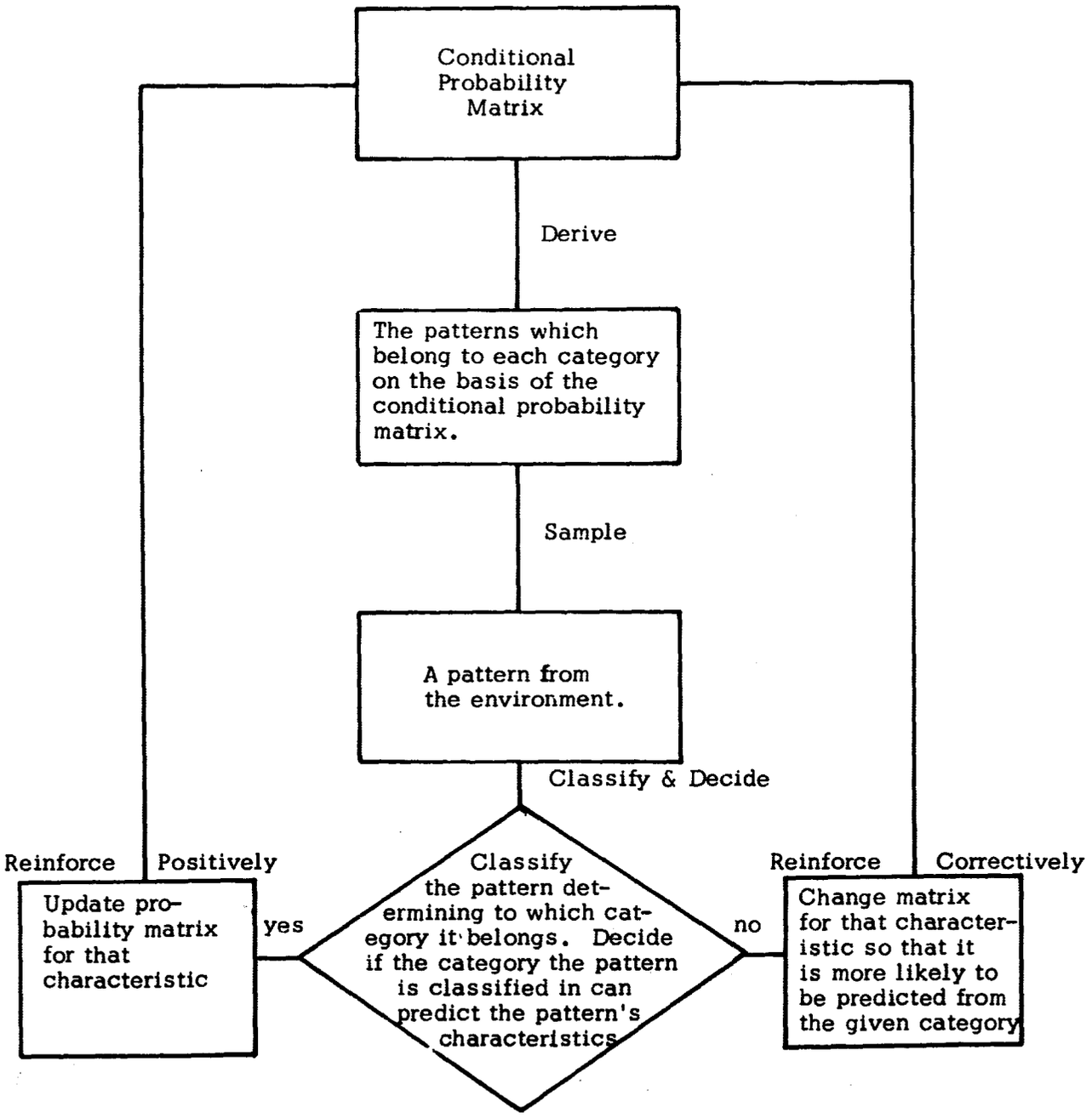

Conditional Probability Matrix
Derive
The patterns which belong to each category on the basis of the conditional probability matrix.
Sample
A pattern from the environment.
Classify & Decide
Classify the pattern determining to which category it belongs. Decide if the category the pattern is classified in can predict the pattern's characteristics
yes
no
Reinforce Positively
Update probability matrix for that characteristic
Reinforce Correctively
Change matrix for that characteristic so that it is more likely to be predicted from the given category

Figure 4.

iately if we think of the storage requirements of the conditional probability of $g = (X_1, X_2, \ldots X_N)$, given category H_i, $P(g/H_i)$, where each X_j is an integer $0 \leq X_j \leq 9$. For N equals one we need ten storage cells, for N equals two we need one hundred, etc.

An analogous problem exists when we try to represent (store) three dimensional objects on two-dimensional paper. However, we usually find we can represent them adequately by drawing six or fewer principal views. Perhaps in a similar manner we could choose a measure, related to conditional probability, where storage requirements would grow linearly with the dimensionality of the patterns.

Let $G = \{g\}$ be the pattern group (finite), $g = (X_1, X_2, \ldots X_D)$, $X_i \epsilon Z = \{L_i\}_{i=1}^{R}$ a finite alphabet or numeric set. G is thus the set of all possible patterns which might result from a measurement. Associated with each pattern g is a probability $P(\{g\})$ where P is the probability distribution over the pattern group G. It is $P(\{g\})$ which has exponentially-growing storage requirements.

Let A be a subset of G; $A \subset G$. Define

$$A_\delta = \begin{cases} A \text{ if and only if } \delta = 1 \\ A^C \text{ if and only if } \delta = -1 \end{cases}$$

Now suppose there exists a set E of subsets E_i of G;

$E = \{E_i | E_i \subset G\}_{i=1}^{N}$ such that for every $g \epsilon G$ there exists a unique sequence $\{\delta_i(g)\}_{i=1}^{N}$ such that $\{g\} = \bigcap_{i=1}^{N} E_{i\delta_i(g)}$. Clearly the order of the set E will grow linearly with increasing dimensionality of g, and storing $P(E_i)$ will not heavily tax available system storage.

The existence of the set E, whose order grows linearly with a linear increase in dimension of g, is assured. We show this by construction.

Let $E = \{E_{ij}\}$, $E_{ij} = \{g \epsilon G | g = (X_1, X_2, \ldots X_D), X_i = L_j\}$, $i = 1, 2, \ldots, D$,

$j = 1,2,\ldots R$. Define for every $g \epsilon G$ $\delta_{ij}(g) =$ $\begin{cases} 1 & \text{if and only if the } i^{th} \text{ coordinate is the } j^{th} \text{ alphabetic or numeric symbol} \\ -1 & \text{otherwise.} \end{cases}$

Clearly $\{g\} = \bigcap_{i=1}^{D} \bigcap_{j=1}^{R} E_{ij\delta_{ij}}(g)$ for every $g \epsilon G$. Let $H = \{H_j\}_{j=1}^{M}$ be the set of similarity sets H_j where $H_j \subset G$.

If $g \epsilon H_k$ then we classify the pattern g as the k^{th} category.

Let $g \epsilon G$. Consider $\dfrac{P(E_{i\delta_i(g)} \cap H_j)}{P(E^c_{i\delta_i(g)} \cap H_j)}$ which is the conditional odds ratio for the i^{th} characteristic, given that g is classified as the j^{th} category. This odds ratio is a measure of the liklihood that g has the i^{th} characteristic against g not having the i^{th} characteristic.

Similarly $\dfrac{P(E^c_{i\delta_i(g)} \cap H^c_j)}{P(E_{i\delta_i(g)} \cap H^c_j)}$ is the conditional odds ratio for not the i^{th} characteristic, given that g is classified as not the j^{th} category. Given that g is classified as not the j^{th} category, this odds ratio is a measure of the liklihood that g does not have the i^{th} characteristic against g having the i^{th} characteristic.

$\dfrac{P(E_{i\delta_i(g)} \cap H_j) P(E^c_{i\delta_i(g)} \cap H^c_j)}{P(E^c_{i\delta_i(g)} \cap H_j) P(E_{i\delta_i(g)} \cap H^c_j)}$ is the product of the odds ratio. We should notice the symmetry between E_i and H_j for this ratio. This implies it is the ratio of the odds ratio for the j^{th} category given the i^{th} characteristic, as well as for the i^{th} characteristic given the j^{th} category.

Considering the meaning of the ratio, we may easily formulate a reasonable decision rule T. We classify a pattern g as the k^{th} category if

and only if

$$\prod_{i=1}^{N} \frac{P(E_{i\delta_i(g)} \cap H_k)\, P(E^c_{i\delta_i(g)} \cap H^c_k)}{P(E^c_{i\delta_i(g)} \cap H_k)\, P(E_{i\delta_i(g)} \cap H^c_k)} = \max_{j=1,2,\ldots M} \prod_{i=1}^{N} \frac{P(E_{i\delta_i(g)} \cap H_j) P(E^c_{i\delta_i(g)} \cap H^c_j)}{P(E^c_{i\delta_i(g)} \cap H_j) P(E_{i\delta_i(g)} \cap H^c_j)}$$

where we choose the smaller k if k is not unique. Define for every $g \in G$, $j=1,\ldots M$,

$$\eta_j(g) = \begin{cases} 1 \text{ if and only if } g \text{ is classified as the } j\text{th category,} \\ -1 \text{ otherwise.} \end{cases}$$

If H is indeed a good partial copy of G, we would expect that, given that g is classified as the k^{th} category, we can predict g's characteristics -- i.e., which E_i's g is in. Thus we want

$$\prod_{j=1}^{M} \frac{P\left(E_{i\delta_i(g)} \cap H_{j\eta_j(g)}\right) P\left(E^c_{i\delta_i(g)} \cap H^c_{j\eta_j(g)}\right)}{P\left(E^c_{i\delta_i(g)} \cap H_{j\eta_j(g)}\right) P\left(E_{i\delta_i(g)} \cap H^c_{j\eta_j(g)}\right)} > y > 1$$

for every i, $i=1,2,\ldots,N$ where y is a given prediction constant. If this is not so we must modify the set H in a way that it is more probable that the inequality will be satisfied next time a pattern is so classified. On the basis of these ideas we now detail an adaptive pattern recognition scheme.

Let $G = \{g\}$ be the pattern group, with probability distribution P. Let A be a subset of G, $A \subset G$. Define $A_\delta = \begin{cases} A \text{ if and only if } \delta = 1 \\ A^c \text{ if and only if } \delta = -1 \end{cases}$

Let $E = \{E_i\}_{i=1}^{N}$ be a set of characteristic subsets of G, such that for every pattern $g \in G$ there exists a unique sequence $\{\delta_i(g)\}_{i=1}^{N}$ such that $\{g\} = \bigcap_{i=1}^{N} E_{i\delta_i(g)}$.

We have already noted the existence of the set E. Thus each pattern may be completely described (represented) by listing its characteristics. If g has the i^{th} characteristic $\delta_i(g)=1$. If g does not have the i^{th} characteristic $\delta_i(g)=-1$. Let M be the number of categories we wish to have.

We define arbitrarily the parameters a^o_{ij}, b^o_{ij}, c^o_{ij}, d^o_{ij}. $i=1,\ldots,N, j=1,\ldots M$

which are the zeroth approximation to $P(E_i \cap H_j)$, $P(E_i^c \cap H_j^c)$, $P(E_i^c \cap H_j)$, $P(E_i \cap H_j^c)$ respectively. H_j is the j^{th} similarity set which we finally wish to construct; we classify g in the j^{th} category if and only if $g \in H_j$.

Let $q_{ij}^o = \dfrac{a_{ij}^o \; b_{ij}^o}{c_{ij}^o \; d_{ij}^o}$, $i=1,\ldots,N$, $j=1,\ldots,M$. Since the parameters are approximations to their respective probabilities, we must have $0 \leq a_{ij}, b_{ij}, c_{ij}, d_{ij} \leq 1$ and $a_{ij}^o + b_{ij}^o + c_{ij}^o + d_{ij}^o = 1$ for every i and j, since the sum of their respective probabilities is 1.

Further, we must have for every j $a_{ij}^o + c_{ij}^o = a_{kj}^o + c_{kj}^o < 1$ for every i and k, since $P(E_i \cap H_j) + P(E_i^c \cap H_j) = P(H_j)$ for every i. Also, for every i, $a_{ij}^o + d_{ij}^o = a_{ik}^o + d_{ik}^o < 1$ for every j and k, since $P(E_i \cap H_j) + P(E_i \cap H_j^c) = P(E_i)$ for every j.

We define the process inductively, describing a complete cycle starting from the n^{th} sample and continuing to sample n + 1. Suppose we have defined the parameters $a_{ij}^n, b_{ij}^n, c_{ij}^n, d_{ij}^n$; $0 \leq a_{ij}^n, b_{ij}^n, c_{ij}^n, d_{ij}^n \leq 1$, $i=1,\ldots,N$, $j=1,\ldots,M$;

$a_{ij}^n + c_{ij}^n = a_{kj}^n + c_{kj}^n < 1$ for every i and k, $j=1,\ldots,M$;

$a_{ij}^n + d_{ij}^n = a_{ik}^n + d_{ik}^n < 1$ for every j and k, $i=1,\ldots,N$.

Let $q_{ij}^n = \dfrac{a_{ij}^n \; b_{ij}^n}{c_{ij}^n \; d_{ij}^n}$ $i=1,\ldots N$, $j=1,\ldots,M$ and let

$$H_k^n = \left\{ g \in G \;\middle|\; \prod_{j=1}^{N} (q_{jk}^n)^{\delta_j(g)} = \max_{m=1,\ldots M} \prod_{j=1}^{N} (q_{jm}^n)^{\delta_j(g)}, \text{ where we take the smaller index k if k is not unique.} \right\}, \quad k = 1,\ldots M.$$

First we should note that $H^n = \{H_k^n\}_{k=1}^{M}$ is a partition of G since by taking the smaller index when k is not unique we have made $H_k^n \cap H_j^n = \emptyset$ for $k \neq j$.

$\bigcup_{k=1}^{M} H_k^n \supset G$ since for every $g \in G$ a maximum product $\prod_{j=1}^{N} (q_{jm}^n)^{\delta_j(g)}$ must exist for some m. $\bigcup_{k=1}^{M} H_k^n \subset G$ since $H_k^n \subset G$ for every k. This implies $\bigcup_{k=1}^{M} H_k^n = G$. Second, we should note that $(q_{jk}^n)^{\delta_j(g)}$ is the n^{th} approximation to $\frac{P(E_{i\delta_i(g)} \cap H_j)\, P(E^c_{i\delta_i(g)} \cap H_j^c)}{P(E^c_{i\delta_i(g)} \cap H_j)\, P(E_{i\delta_i(g)} \cap H_j^c)}$. This is seen easily by considering

$$\frac{P(E_{i\delta_i(g)} \cap H_j)\, P(E^c_{i\delta_i(g)} \cap H_j^c)}{P(E^c_{i\delta_i(g)} \cap H_j)\, P(E_{i\delta_i(g)} \cap H_j^c)} = \begin{cases} \dfrac{P(E_i \cap H_j)\, P(E_i^c \cap H_j^c)}{P(E_i^c \cap H_j)\, P(E_i \cap H_j^c)} & \text{when } \delta_i(g) = 1 \\ \dfrac{P(E_i^c \cap H_j)\, P(E_i \cap H_j^c)}{P(E_i \cap H_j)\, P(E_i^c \cap H_j^c)} & \text{when } \delta_i(g) = -1. \end{cases}$$

We sample a pattern g_{n+1} from G and classify it in the k^{th} category, where k is the smallest index m which maximizes $\prod_{j=1}^{N} (q_{jm}^n)^{\delta_j(g_{n+1})}$. Given that g_{n+1} has been classified in the k^{th} category, we try to predict the characteristics for g_{n+1}. We will positively reinforce when half or more of the predictors $\hat{\delta}_i(g_{n+1})$ are correct predictions, and negatively reinforce otherwise.

Let an epsilon between zero and one be given, $0<\epsilon<1$, such that $1/\epsilon$ represents approximately the number of patterns over which we expect the probability distribution to be stationary. The reinforcement is such that the effects of early patterns (those earlier than $1/\epsilon$ from the present) tend to be negligible.

Since g_{n+1} has been classified in the k^{th} category, we define $\eta_k(g_{n+1})=1$ and $\eta_j(g_{n+1})=-1$, $j \neq k$, $j=1,\ldots,M$. Given the k^{th} category, we now try to predict

the characteristics for g_{n+1}. We predict for i=1,...,N

$$\hat{\delta}_i(g_{n+1}) = \begin{cases} 1 \text{ if } \prod_{j=1}^{M} (q_{ij}^n)^{\eta_j(g_{n+1})} > y \geq 1 \\ -1 \text{ otherwise.} \end{cases}$$

y is a chosen prediction constant greater than one.

Once a pattern has been classified and the sequence $\{\hat{\delta}_i(g_{n+1})\}$ has been determined, there are eight possibilities for each i and j:

1) $g_{n+1} \epsilon E_i \cap H_j^n$ and $\hat{\delta}_j(g_{n+1}) = \delta_i(g_{n+1})$,

2) $g_{n+1} \epsilon E_i \cap H_j^n$ and $\hat{\delta}_i(g_{n+1}) \neq \delta_i(g_{n+1})$,

3) $g_{n+1} \epsilon E_i^c \cap H_j^{nc}$ and $\hat{\delta}_i(g_{n+1}) = \delta_i(g_{n+1})$,

4) $g_{n+1} \epsilon E_i^c \cap H_j^{nc}$ and $\hat{\delta}_i(g_{n+1}) \neq \delta_i(g_{n+1})$,

5) $g_{n+1} \epsilon E_i^c \cap H_j^n$ and $\hat{\delta}_i(g_{n+1}) = \delta_i(g_{n+1})$,

6) $g_{n+1} \epsilon E_i^c \cap H_j^n$ and $\hat{\delta}_i(g_{n+1}) \neq \delta_i(g_{n+1})$,

7) $g_{n+1} \epsilon E_i \cap H_j^{nc}$ and $\hat{\delta}_i(g_{n+1}) = \delta_i(g_{n+1})$,

8) $g_{n+1} \epsilon E_i \cap H_j^{nc}$ and $\hat{\varepsilon}_i(g_{n+1}) \neq \delta_i(g_{n+1})$.

For each of the possibilities we indicate the reinforcement which satisfies the following conditions for the n+1 iteration given, that the parameters satisfy these conditions for the n^{th} iteration.

For all i and j:

1) $0 \leq a_{ij}^{n+1}, b_{ij}^{n+1}, c_{ij}^{n+1}, d_{ij}^{n+1} \leq 1$,

2) $a_{ij}^{n+1} + b_{ij}^{n+1} + c_{ij}^{n+1} + d_{ij}^{n+1} = 1$,

3) $a_{ij}^{n+1} + c_{ij}^{n+1} = a_{kj}^{n+1} + c_{kj}^{n+1} < 1$ for all k,

4) $a_{ij}^{n+1} + d_{ij}^{n+1} = a_{ik}^{n+1} + d_{ik}^{n+1} < 1$ for all k.

ϵ is chosen so that $1/\epsilon$ is approximately the number of patterns over which the probability distribution of G is stationary. ϵ^* may be a few orders of magnitude higher than ϵ. Positive reinforcement is used when

$$\Sigma\left|\delta_i(g_{n+1}) - \delta_i(g_{n+1})\right| \leq N$$

Negative reinforcement is used otherwise.

Positive Reinforcement Scheme

If $g_{n+1} \epsilon E_i \cap H_j^n$ then:

$$a_{ij}^{n+1} = a_{ij}^n (1-\epsilon) + \epsilon$$

$$b_{ij}^{n+1} = b_{ij}^n (1-\epsilon)$$

$$c_{ij}^{n+1} = c_{ij}^n (1-\epsilon)$$

$$d_{ij}^{n+1} = d_{ij}^n (1-\epsilon)$$

If $g_{n+1} \epsilon E_i^c \cap H_j^{nc}$ then:

$$a_{ij}^{n+1} = a_{ij}^n (1-\epsilon)$$

$$b_{ij}^{n+1} = a_{ij}^n (1-\epsilon) + \epsilon$$

$$c_{ij}^{n+1} = c_{ij}^n (1-\epsilon)$$

$$d_{ij}^{n+1} = d_{ij}^n (1-\epsilon)$$

If $g_{n+1} \epsilon E_i \cap H_j^{nc}$ then:

$$a_{ij}^{n+1} = a_{ij}^n (1-\epsilon)$$

$$b_{ij}^{n+1} = b_{ij}^n (1-\epsilon)$$

$$c_{ij}^{n+1} = c_{ij}^n (1-\epsilon)$$

$$d_{ij}^{n+1} = d_{ij}^n (1-\epsilon) + \epsilon$$

If $g_{n+1} \epsilon E_i^c \bigcap H_j^n$ then:

$$a_{ij}^{n+1} = a_{ij}^n (1-\epsilon)$$

$$b_{ij}^{n+1} = b_{ij}^n (1-\epsilon)$$

$$c_{ij}^{n+1} = c_{ij}^n (1-\epsilon) + \epsilon$$

$$d_{ij}^{n+1} = d_{ij}^n (1-\epsilon)$$

Negative Reinforcement Scheme

If $g_{n+1} \epsilon E_i \bigcap H_j^n$ or $g_{n+1} \epsilon E_i^c \bigcap H_j^{nc}$ then:

$$a_{ij}^{n+1} = a_{ij}^n (1 + \epsilon^* b_{ij}^n c_{ij}^n d_{ij}^n)$$

$$b_{ij}^{n+1} = b_{ij}^n (1 + \epsilon^* a_{ij}^n c_{ij}^n d_{ij}^n)$$

$$c_{ij}^{n+1} = c_{ij}^n (1 - \epsilon^* a_{ij}^n b_{ij}^n d_{ij}^n)$$

$$d_{ij}^{n+1} = d_{ij}^n (1 - \epsilon^* a_{ij}^n b_{ij}^n c_{ij}^n)$$

If $g_{n+1} \epsilon E_i \bigcap H_j^{nc}$ or $g_{n+1} \epsilon E_i^c \bigcap H_j^n$ then:

$$a_{ij}^{n+1} = a_{ij}^n (1 - \epsilon^* b_{ij}^n c_{ij}^n d_{ij}^n)$$

$$b_{ij}^{n+1} = b_{ij}^n (1 - \epsilon^* a_{ij}^n c_{ij}^n d_{ij}^n)$$

$$c_{ij}^{n+1} = c_{ij}^n (1 + \epsilon^* a_{ij}^n b_{ij}^n d_{ij}^n)$$

$$d_{ij}^{n+1} = d_{ij}^n (1 + \epsilon^* a_{ij}^n b_{ij}^n c_{ij}^n)$$

The motivation for positive reinforcement is illustrated by the following example. Suppose we are determining the probability of rain. Each day we check to see if it rains. By the N^{th} day the probability that it rains may be expressed as:

$$P_N(\text{rain}) = \frac{\text{number of days it rained}}{\text{total number of days observed}} = \frac{M}{N}$$

Suppose that it rains on the $N+1^{st}$ day.

$$P_{N+1}(\text{rain}) = \frac{M+1}{N+1} = \frac{NP_N(\text{rain}) + 1}{N+1} = \frac{N}{N+1} P_N(\text{rain}) + \frac{1}{N+1} .$$

Suppose that it does not rain on the $N+1^{st}$ day. Then

$$P_{N+1}(\text{rain}) = \frac{M}{N+1} = \frac{N}{N+1} P_N(\text{rain}) . \text{ If we let } \epsilon = \frac{1}{N+1} \text{ then } \frac{N}{N+1} = 1-\epsilon .$$

Thus if it rains on the $N+1^{st}$ day, $P_{N+1}(\text{rain}) = P_N(\text{rain})\ (1-\epsilon) + \epsilon$, and if it does not rain, then $P_{N+1}(\text{rain}) = P_N(\text{rain})\ (1-\epsilon)$. This is the same format as the positive-reinforcement scheme. We should note that in each case this reinforcement satisfies the conditions on page 18.

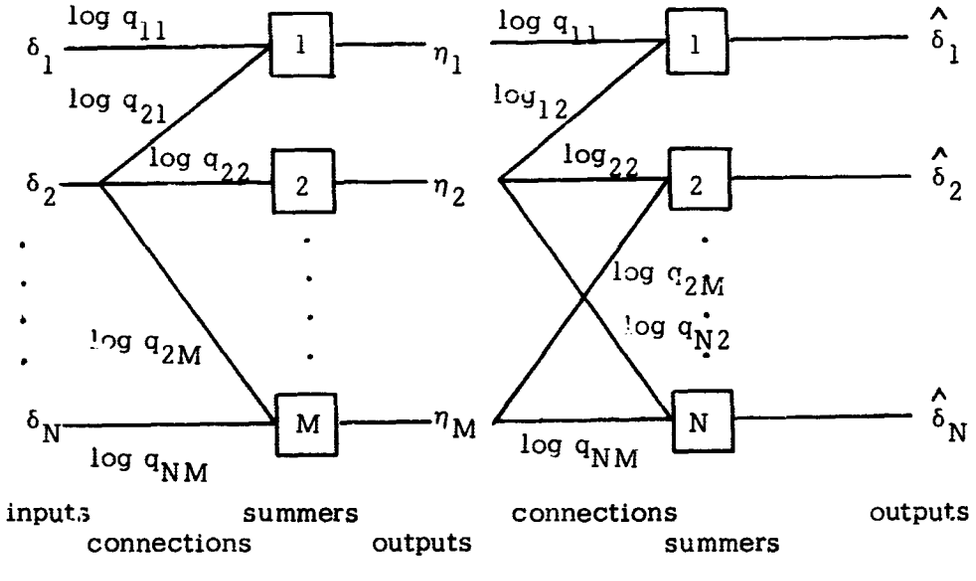
The motivation for the negative reinforcement is as follows. Suppose $\sum_{i=1}^{N} \left| \hat{\delta}_i(g_{n+1}) - \delta_i(g_{n+1}) \right| > N$. This means that more than half of the predictors $\hat{\delta}_i(g_{n+1})$ predict wrong. Thus, we wish to change the parameters a_{ij}, b_{ij}, c_{ij}, d_{ij} so that more of the predictors $\hat{\delta}_i(g_{n+1})$ predict right. Suppose $g_{n+1} \epsilon E_i \cap H_j^n$. Since $\hat{\delta}_i$ is the predictor for δ_i we want $\hat{\delta}_i = \delta_i = 1$, which implies that we want $\prod_{k=1}^{M} q_{ik}^{n_k}(g_{n+1}) > 1$.

We can negatively or correctively reinforce by increasing each term $(q_{ik})^{n_k}$ in the product. Since $g_{n+1} \epsilon E_i \cap H_j^n$ we know $n_j = 1$ and $n_k = -1$ for $k \neq j$. Therefore, when $k \neq j$, we increase c_{ik} and d_{ik} and decrease a_{ik} and b_{ik}. When $k = j$, we increase a_{ik} and b_{ik} and decrease c_{ik} and d_{ik}. The reasoning is similar for the other cases, and we should note that in each case the negative reinforcement satisfies the four conditions

on page 18: in addition, it was chosen so that the $n+1^{st}$ approximation to $P(H_j)$ and $P(E_i)$ remains the same as the n^{th} approximation. Thus $a_{ij}^{n+1} + c_{ij}^{n+1} = a_{ij}^{n} + c_{ij}^{n}$ and $a_{ij}^{n+1} + d_{ij}^{n+1} = a_{ij}^{n} + d_{ij}^{n}$. Only the relative weight between a_{ij} and c_{ij} and between a_{ij} and d_{ij} is changed so that it is more likely for δ_i to be correct in the next iteration.

Similarity to Perceptrons

We should also note that the adaptive method described here is similar in structure to many two layer perceptron schemes. Each associative unit is a summer which responds to an input vector by way of a matrix of weighted connections $Q = \{q_{ij}\}$, where q_{ij} is the connection from the i^{th} component of the input vector to the j^{th} associative unit.



Perceptron Representation of Adaptive Machine

Now, using vector notation, we may easily describe the operation of the adaptive machine. First we define two vector operators. Let V be a vector,

$V = \begin{bmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_N \end{bmatrix}$. Define: $\text{Amax } V = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_N \end{bmatrix}$ where $a_k = 1$ if and only if k is the smallest index such that $v_k = \max_{i,i=1,\ldots N} v_i$ and all other a_i's are -1.

$\text{Sgn } V = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_N \end{bmatrix}$ where $b_i = 1$ if and only if $v_i > 0$ and $b_i = -1$ otherwise.

Let the input to the first associative layer be $\Delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_N \end{bmatrix}$ and the output be $H = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_M \end{bmatrix}$. If A is a matrix, then A^T is A transpose. Therefore we have $H = A \max \{\log Q^T\} \Delta$ and $\hat{\Delta} = \begin{bmatrix} \hat{\delta}_1 \\ \hat{\delta}_2 \\ \hat{\delta}_N \end{bmatrix} = \text{Sgn}\{(\log Q)\, H\}$.

A further simplification results if we logarithmically transform all parameters and the reinforcement schemes as follows. We might start as before and arbitrarily define a_{ij}^o, b_{ij}^o, c_{ij}^o, d_{ij}^o so that

1) $e^{a_{ij}^o} + e^{b_{ij}^o} + e^{c_{ij}^o} + e^{d_{ij}^o} = 1$ and

2) $e^{a_{ij}^o} + e^{c_{ij}^o} = e^{a_{kj}^o} + e^{c_{kj}^o}$ for all k

3) $e^{a_{ij}^o} + e^{d_{ij}^o} = e^{a_{ik}^o} + e^{d_{ik}^o}$ for all k.

These three conditions are the transformed last three conditions on page 18. The first condition on page 18 does not need to be transformed, since e raised to any real finite power is positive.

Next we must determine the transformed reinforcement schemes. Consider the positive reinforcement scheme which has as a basic element the equation $x^{n+1} = x^{n+1}(1-\epsilon) + \epsilon$. This implies

$$\log x^{n+1} = \log x^n + \log \left(1-\epsilon + \frac{\epsilon}{x^n}\right)$$

$$= \log x^n + \log \left[(1-\epsilon)\left(1 + \frac{\epsilon}{(1-\epsilon)x^n}\right)\right]$$

$$= \log x^n + \log (1-\epsilon) + \log \left[\left(1 + \frac{\epsilon}{(1-\epsilon)x^n}\right)\right]$$

Since $\epsilon \ll 1$, we may approximate $\log (1-\epsilon) \approx -\epsilon$ and $\log \left(1 + \frac{\epsilon}{(1-\epsilon)x^n}\right) \approx \frac{\epsilon}{(1-\epsilon)x^n} \approx \frac{\epsilon}{x^n}$. If we let $y^n = \log x^n$ and $y^{n+1} = \log x^{n+1}$, the transformed equation becomes $y^{n+1} = y^n - \epsilon + \epsilon e^{-y^n}$. Obviously the equation $x^{n+1} = x^n (1-\epsilon)$ will be transformed to $y^{n+1} = y^n - \epsilon$.

Positive Reinforcement Scheme:

$$\sum_{i=1}^{N} \left| \delta_i (g_{n+1}) - \hat{\delta}_i (g_{n+1}) \right| \leq N$$

and ϵ is chosen so that $1/\epsilon$ is approximately the number of consecutive patterns in the sample over which the probability distribution is expected to be stationary.

If $g_{n+1} \in E_i \cap H_j^n$ then:

$$a_{ij}^{n+1} = a_{ij}^n - \epsilon (1 - e^{-a_{ij}^n})$$

$$b_{ij}^{n+1} = b_{ij}^n - \epsilon$$

$$c_{ij}^{n+1} = c_{ij}^n - \epsilon$$

$$d_{ij}^{n+1} = d_{ij}^n - \epsilon$$

If $g_{n+1} \in E_i^c \cap H_j^{nc}$ then:

$$a_{ij}^{n+1} = a_{ij}^n - \epsilon$$

$$b_{ij}^{n+1} = b_{ij}^n - \epsilon (1 - e^{-b_{ij}^n})$$

$$c_{ij}^{n+1} = c_{ij}^n - \epsilon$$

$$d_{ij}^{n+1} = d_{ij}^n - \epsilon$$

If $g_{n+1} \in E_i^c \cap H_j^n$ then:

$$a_{ij}^{n+1} = a_{ij}^n - \epsilon$$

$$b_{ij}^{n+1} = b_{ij}^n - \epsilon$$

$$c_{ij}^{n+1} = c_{ij}^n - \epsilon (1 - e^{-c_{ij}^n})$$

$$d_{ij}^{n+1} = d_{ij}^n - \epsilon$$

Now consider the negative reinforcement scheme. Here equations such as $a^{n+1} = a^n (1 + \epsilon^* b^n c^n d^n)$ must be transformed. This implies $\log a^{n+1} = \log a^n + \log 1 + \epsilon^* b^n c^n d^n$. Since $\epsilon^* b^n c^n d^n < 1$, we may approximate: $\log a^{n+1} \approx \log a^n + \epsilon^* b^n c^n d^n$. If we let $x^{n+1} = \log a^{n+1}$, $u^n = \log b^n$, $v^n = \log c^n$, $y^n = \log d^n$, the transformed equation becomes $x^{n+1} = x^n + \epsilon^* e^{u^n + v^n + y^n}$.

The entire transformed adaptive system may easily be described. We define the process inductively, starting from the n^{th} iteration and continuing to iteration n+1. Suppose we have defined the parameters a_{ij}^n, b_{ij}^n, c_{ij}^n, d_{ij}^n such that:

1) $e^{a_{ij}^n} + e^{b_{ij}^n} + e^{c_{ij}^n} + e^{d_{ij}^n} = 1$, $i=1,\ldots N$, $j=1,\ldots M$.

2) $e^{a_{ij}^n} + e^{c_{ij}^n} = e^{d_{kj}^n} + e^{c_{kj}^n}$ for every i and k, $j=1,\ldots M$.

3) $e^{a_{ij}^n} + e^{d_{ij}^n} = e^{a_{ik}^n} + e^{d_{ik}^n}$ for every j and k, $i = 1,\ldots N$.

Let $q_{ij}^n = a_{ij}^n + b_{ij}^n - c_{ij}^n - d_{ij}^n$, $i = 1,\ldots N$, $j = 1,\ldots M$ and let

$$H_k^n = \left\{ g \epsilon G \left| \sum_{j=1}^{N} (q_{jk}^n) \delta_j(g) = \max_{m=1,\ldots M} \sum (q_{jm}^n) \delta_i(g) \right. \right\} \quad \begin{array}{l} \text{where we take the} \\ \text{smaller index k if} \\ \text{k is not unique} \end{array} \quad k = 1,\ldots M.$$

Thus we sample a pattern g_{n+1} from G and classify it in the k^{th} category where k is the smallest index m which maximizes $(q_{jm}^n) \delta_j(g_{n+1})$. Given that g_{n+1} has been classified in the k^{th} category, we try to predict the characteristics for g_{n+1}. Since g_{n+1} has been classified in the k^{th} category we define $\eta_k(g_{n+1}) = 1$ and $\eta_j(g_{n+1}) = -1$, $j \neq k$, $j= 1,\ldots M$. Given the k^{th} category, we now try to predict the characteristics for g_{n+1}. We predict for $i = 1,\ldots N$ $\delta_i^n(g_{n+1}) = \begin{cases} 1 \text{ if } \sum_{j=1}^{M} (q_{ij}^n) \eta_j(g_{n+1}) > y \geq 0 \\ -1 \text{ otherwise.} \end{cases}$

y is a chosen prediction constant greater than one.

If $g_{n+1} \in E_i \cap H_j^{nc}$ then:

$$a_{ij}^{n+1} = a_{ij}^{n} - \epsilon$$

$$b_{ij}^{n+1} = b_{ij}^{n} - \epsilon$$

$$c_{ij}^{n+1} = c_{ij}^{n} - \epsilon$$

$$d_{ij}^{n+1} = d_{ij}^{n} - \epsilon \ (1 - e^{-d_{ij}^{n}})$$

Negative Reinforcement Scheme:

$$\sum_{i=1}^{N} \left| \delta_i \ (g_{n+1}) - \hat{\delta}_i \ (g_{n+1}) \right| > N$$

and ϵ^* is a few orders of magnitude higher than ϵ .

If $g_{n+1} \in E_i \cap H_j^{n}$ or $g_{n+1} \in E_i^{c} \cap H_j^{nc}$ then:

$$a_{ij}^{n+1} = a_{ij}^{n} + \epsilon^* \, e^{b_{ij}^{n}+c_{ij}^{n}+d_{ij}^{n}}$$

$$b_{ij}^{n+1} = b_{ij}^{n} + \epsilon^* \, e^{a_{ij}^{n}+c_{ij}^{n}+d_{ij}^{n}}$$

$$c_{ij}^{n+1} = c_{ij}^{n} - \epsilon^* \, e^{a_{ij}^{n}+b_{ij}^{n}+d_{ij}^{n}}$$

$$d_{ij}^{n+1} = d_{ij}^{n} - \epsilon^* \, e^{a_{ij}^{n}+b_{ij}^{n}+c_{ij}^{n}}$$

If $g_{n+1} \in E_i^{c} \cap H_j^{n}$ or $g_{n+1} \in E_i \cap H_j^{nc}$ then:

$$a_{ij}^{n+1} = a_{ij}^{n} - \epsilon^* \, e^{b_{ij}^{n}+c_{ij}^{n}+d_{ij}^{n}}$$

$$b_{ij}^{n+1} = b_{ij}^{n} - \epsilon^* \, e^{a_{ij}^{n}+c_{ij}^{n}+d_{ij}^{n}}$$

$$c_{ij}^{n+1} = c_{ij}^{n} + \epsilon^* \, e^{a_{ij}^{n}+b_{ij}^{n}+d_{ij}^{n}}$$

$$d_{ij}^{n+1} = d_{ij}^{n} + \epsilon^* \, e^{a_{ij}^{n}+b_{ij}^{n}+c_{ij}^{n}}$$

Of course $q_{ij}^{n+1} = a_{ij}^{n+1} + b_{ij}^{n+1} - c_{ij}^{n+1} - d_{ij}^{n+1}$

We may now summarize both transformed reinforcement schemes in the following table.

$$a_{ij}^{n+1} = a_{ij}^{n} + \epsilon \frac{|\delta_i - \hat{\delta}_i|}{2} \left\{ \frac{(1+\eta_j)(1+\delta_i)}{4} e^{-a_{ij}^{n}} - 1 \right\} + \frac{|\delta_i + \delta_i|}{2} \eta_j \delta_i \epsilon^* e^{b_{ij}^{n} + c_{ij}^{n} + d_{ij}^{n}}$$

$$b_{ij}^{n+1} = b_{ij}^{n} + \epsilon \frac{|\delta_i - \delta_i|}{2} \left\{ \frac{(1-\eta_j)(1-\delta_i)}{4} e^{-b_{ij}^{n}} - 1 \right\} + \frac{|\delta_i + \delta_i|}{2} \eta_j \delta_i \epsilon^* e^{a_{ij}^{n} + c_{ij}^{n} + d_{ij}^{n}}$$

$$c_{ij}^{n+1} = c_{ij}^{n} + \epsilon \frac{|\delta_i - \delta_i|}{2} \left\{ \frac{(1+\eta_j)(1-\delta_i)}{4} e^{-c_{ij}^{n}} - 1 \right\} - \frac{|\delta_i + \delta_i|}{2} \eta_j \delta_i \epsilon^* e^{a_{ij}^{n} + b_{ij}^{n} + d_{ij}^{n}}$$

$$d_{ij}^{n+1} = d_{ij}^{n} + \epsilon \frac{|\delta_i - \delta_i|}{2} \left\{ \frac{(1-\eta_j)(1+\delta_i)}{4} e^{-d_{ij}^{n}} - 1 \right\} - \frac{|\delta_i + \delta_i|}{2} \eta_j \delta_i \epsilon^* e^{a_{ij}^{n} + b_{ij}^{n} + c_{ij}^{n}}$$

The transformed matrix equations would be:

$$H = A \max \left[Q^T \Delta \right]$$

$$\hat{\Delta} = \mathrm{Sgn} \left[Q H \right]$$