AD-A039 071 FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OHIO F/G 9/2
A NON-PARAMETRIC ALGORITHM FOR PATTERN CLASSIFICATION, (U)
NOV 76 W MALINA

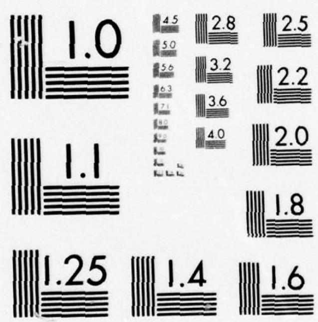
UNCLASSIFIED FTD-ID(RS)I-1209-76 NL

| OF |

AD
A039071

END

DATE
FILMED
5 — 77



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

FTD-ID(RS)I-1209-76

AD-A039071

1

FOREIGN TECHNOLOGY DIVISION

AIR FORCE SYSTEMS COMMAND

A NON-PARAMETRIC ALGORITHM FOR PATTERN CLASSIFICATION

by

Witold Malina

FOREIGN TECHNOLOGY DIVISION

Approved for public release;
distribution unlimited.

FTD ID(RS)I-1209-76

EDITED TRANSLATION

FTD-ID(RS)I-1209-76 5 November 1976

FTD-76-C-001139

A NON-PARAMETRIC ALGORITHM FOR PATTERN CLASSIFICATION

By: Witold Malina

English pages: 30

Source: Archiwum Automatyki i Telemechaniki,
Vol 20, Nr 4, 1975, PP. 483-501.

Country of origin: POLAND
Translated by: SCITRAN
F33657-76-D-0390
Requester: FTD/ETCK

THIS TRANSLATION IS A RENDITION OF THE ORIGINAL FOREIGN TEXT WITHOUT ANY ANALYTICAL OR EDITORIAL COMMENT. STATEMENTS OR THEORIES ADVOCATED OR IMPLIED ARE THOSE OF THE SOURCE AND DO NOT NECESSARILY REFLECT THE POSITION OR OPINION OF THE FOREIGN TECHNOLOGY DIVISION.	PREPARED BY: TRANSLATION DIVISION FOREIGN TECHNOLOGY DIVISION WP-AFB, OHIO.

A NON-PARAMETRIC ALGORITHM FOR PATTERN CLASSIFICATION by Witold Malina

The classification algorithm described can be used in the case of /483 incomplete initial information. A preprocessing layer, a transformation layer and a decision layer are distinguished in the classifier from the functional point of view. During study a rule was applied according to which individual classifier layers were later adjusted.

The essence of the algorithm concerns the division of pattern space into cells. Then a histogram is devised for each class, an estimate of the unknown distribution of the probability of a pattern arising from particular cells. The construction of cells and histograms takes the form of initial information into consideration. The details of the algorithm are explained by example.

Part Two gives an idea of the practical implementation of the basic unit of the classifier in the form of two layers of adaptive logic networks. Each layer of the logical network can be composed of ASL-3 modules. The ASL-3 module (Figure 9) is the basic structural unit of the classifier. It is an adaptive logic network with three inputs and one output. When a classifier is built, there is a possibility of connecting it in a simple way for work with a computer.

I. Description of the Algorithm

1. General Considerations and Assumptions

The basic part of the classifier consists of two adaptive logic network layers. In connection with this, all input, interlayer and output signals can have the form of binary pulses, in order to enable classification of patterns even when input signals y_1, y_2, ... y_N are not binary or when their

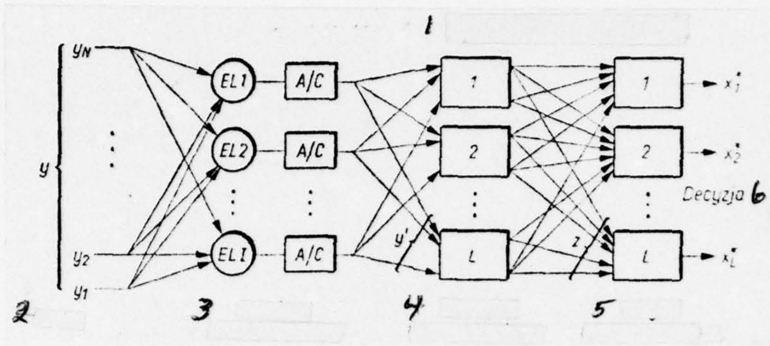


Figure 1. Structure of logic classifier. Key: 1, basic part (logic), 2-pattern, 3-initial processing layer, 4-first logic layer, 5-second logic layer. 6 - decision

number exceeds the input number of the basic part of the classifier, a layer for preprocessing, composed of linear elements EL and analog-digital A/C [AD] converters, was introduced. On the structural side the number of dimensions of the original pattern space is limited by the number N of the coeffic ents weights of linear elements. From the functional point of view we can distinguish the following elements in the structure of the logic part of the classifier (network):

a) Vertically, two layers which fulfill the functions of transformation and decision register respectively, and

b) Horizontally, L, identical layers corresponding to discriminators in other classification systems. A general block diagram of the logic layer classifier for L classes is shown in Figure 1. /484

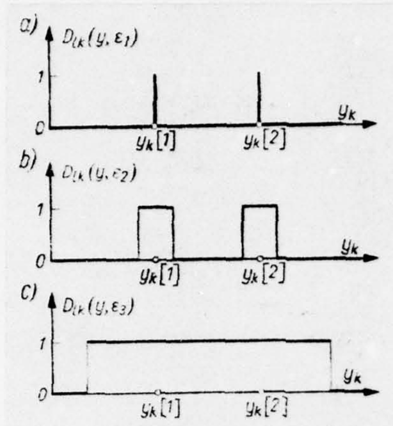


Figure 2. Examples of the route of auxiliary functions for:

a) $\varepsilon_1 = 1$, b) $\varepsilon_2 < 1$, c) $\varepsilon_3 < \varepsilon_2$

The assumptions,on which the working algorithm of the system discussed is based, are divided into two groups.

A. Assumptions About Patterns

1) There exists an absolute N-dimensional space of features $y_1 \ldots y_N$ describing the properties of classified patterns. The numerical values of their parameters can be discrete (binary in particular) or variable in a steady manner in determined sections.

2) As a result of pattern description by the absolute number of parameters and of their quantification, the correspondence between pattern and numerical values describing parameters is not ambiguous.

3) The statistical properties of patterns and of parameters describing them are unknown; observations may have a multimodal distribution.

4) An acquaintance with class x_l is based on given sets of observation $Y(x_l)$, $l = 1, 2, \ldots L$.

B. Assumptions Referring to the Algorithm

1) Classifier decisions must have the form of a binary route.

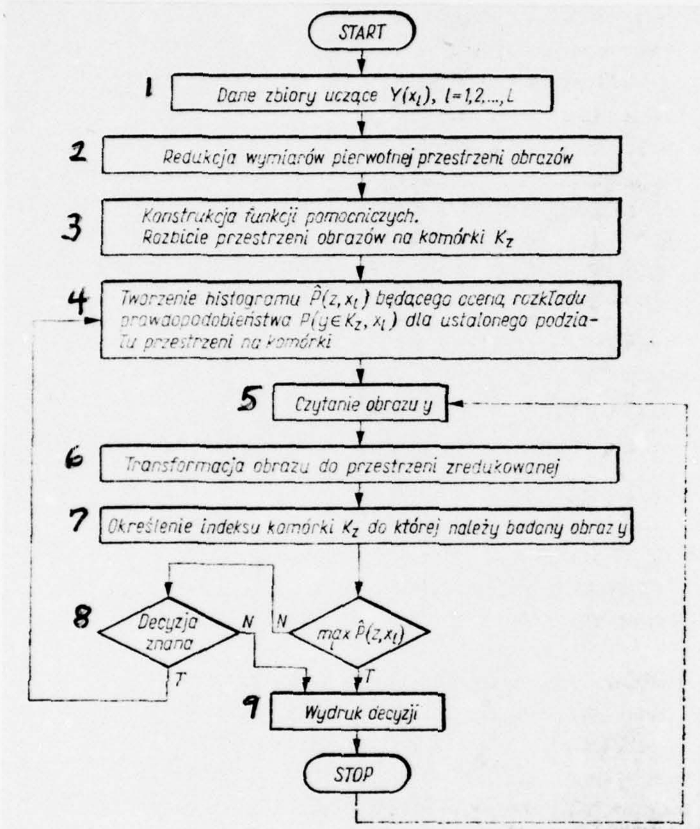


Figure 3. Simplified block diagram of full study$_{\wedge}^{4}$classification algorithm. Key: 1-data set studying $Y(x_1)$, $l = 1, 2, \ldots, L$, 2-reduction of dimensions of original pattern space, 3-construction of auxiliary functions. Breakdown of pattern space into cells K_z, 4-creation of histogram $\hat{P}(z, x_l)$, an evaluation of the distribution of probability $P(y \in K_z, x_l)$ for a determined division of space into cells, 5-reading of pattern y, 6-transformation of pattern to reduced space, 7-definition of cell index K_z to which pattern y investigated belongs, 8-known decision, 9-desision printout.

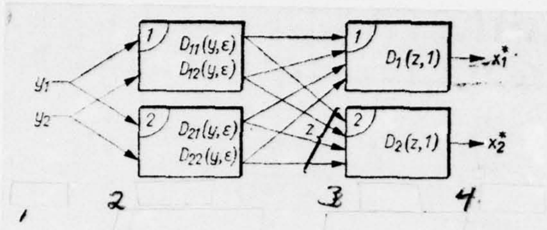


Figure 4. Structure of bidimensional pattern classifier. Key: 1-pattern, 2-first logic layer, 3-second logic layer, 4-decision.

2) It is assumed that the decisive rule has the form

$$x^*(y) = x_l, \text{ if } P(y \in K_z, x_l) > P(y \in K_z, x_j), j = 1, 2 \dots L, j \neq l. \tag{1}$$

where K_z is the z cell (section) occurring as a result of the division of pattern space by hyperplanes.

3) The main probability of classification error

$$\bar{P} = \sum_z P(z)[1 - \max_l P(x_l|z)] = \sum_z \Big[\sum_{l=1}^{L} P(z, x_l) - \max_l P(z, x_l)\Big], \tag{2}$$

is adopted as a criterion for the quality of the decisive rule,

Where: P(z) is the probability of an observation of y occurring in the z cell:

$P(x_l/z)$ is the *aposteriori* probability that an observation belonging to the cell K_z is an element of class x_l;

$P(z,x_l)$ is the total probability that y belongs to a cell of index z and is an element of class x_l.

In general the probabilities occurring above are not known. We shall estimate them in a study cycle on the basis of a study set. /485

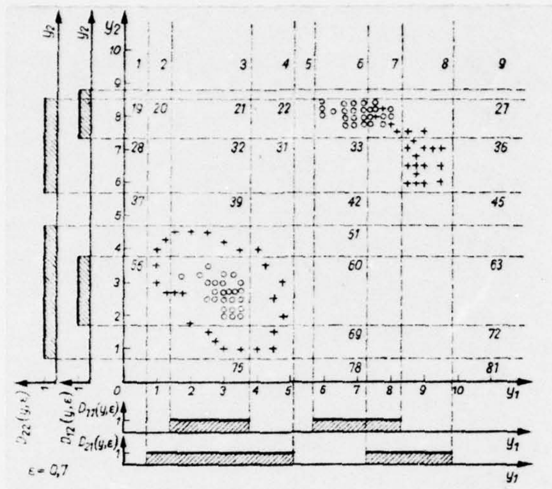


Figure 5. Illustration of method of dividing pattern space into elementary cells with the aid of auxiliary functions.

In addition the algorithm should consider aspects of structural realization with attention to the possibilities of:

a) construction of an independent system,

b) cooperation with any electronic computer.

2. Study and Classification Algorithm

The classifier study cycle is based on the given sets $Y(x_l)$, $l = 1;. 2,\ldots,L$, and is composed of two stages:

-Adjustment of the preprocessing layer,

-choice of the structure of two logic network layers.

In the study cycle we made use of a concept of successive adaptation of individual classifier layers. A detailed description of the adaptation of each layer will be given below.

2.1. Adaptation of Initial Processing Layer

A need for the existence of a preprocessing layer occurs when the number of vector components of the pattern exceeds the input number I

Table 1. Comparison of z routes and elementary cells corresponding to them.

z'	z	l
0	0000	1, 5, 9, 37, 41, 45, 73, 77, 81
1	0001	6, 42, 78
2	0010	10, 14, 18, 44
3	0011	15
4	0100	2, 4, 8, 38, 40, 74, 76
5	0101	3, 7, 39, 43, 75, 79
6	0110	11, 13
7	0111	12, 16
8	1000	28, 32, 36, 50, 54, 64, 68, 72
9	1001	33, 51, 69
10	1010	19, 23, 27, 52, 55, 59, 63
11	1011	24, 60
12	1100	29, 31, 35, 47, 49, 53. 65, 67, 71
13	1101	22, 30, 34, 48, 66, 70
14	1110	20, 26, 56, 58, 62
15	1111	21, 25, 57, 61

Key: l-elementary cells entering composition K_z.

of the logic part of the classifier or when the signals describing the pattern are not binary. In other cases we can omit the initial layer, and introduce the pattern signals directly into the logic part of the classifier.

That space obtained as a result of the transformation made by the initial layer we shall call reduced space. The coordinates of this space are identified successively on the basis of research on the mutual overlap of study sets $Y(x_l)$, $l=1, 2 \ldots L$.

The overlapping of the sets is described by the matrix $[\Delta]$, of which the elements Δ_{ij} are a numerical estimate of the magnitude of covering of sets $Y(x_i)$ and $Y(x_j)$[1]. A decision in regard to a successive, additional k coordinate of reduced space is made on the basis of examination of the breakdown of the elements of the sets studied along the coordinates mentioned earlier.

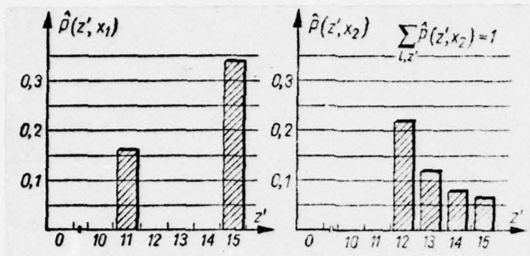


Figure 6. ~~Figure 6~~. Histograms of frequency of observations occurring from classes x_1 and x_2. Parameter ε = 0.7.

As a result of this examination the elements Δ_{ij}, for which the relations

$$\Delta_{ij} \leqslant \Delta_0, \quad i \neq j, \tag{3}$$

where Δ_0 is the given threshold value, is fulfilled, are replaced by zeroes in the matrix $[\Delta]$. The algorithm completes the search for new directions, if all the elements of the matrix $[\Delta]$ were made equal to zero(then $k=k_0$; $k_0 < I$) or $k=I$.

Finally the transformation made by the initial value of N-dimensional space in the reduced k_0-dimensional space can be written in a matrix form

$$\begin{bmatrix} y'_1 \\ \vdots \\ y'_{k_0} \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1N} \\ & \dots & \\ a_{k_0 1} & \dots & a_{k_0 N} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \tag{4}$$

where: a_{kj} is an element of the matrix transformation sought, while (a_{k1}, 1486
$a_{k2}, \dots, a_{kN}$) are a set of coefficients of the linear k element;
y'_k is the k component of the vector of pattern y after transformation;
$k=1,2,\dots,k_0, k_0 \leqslant I$.

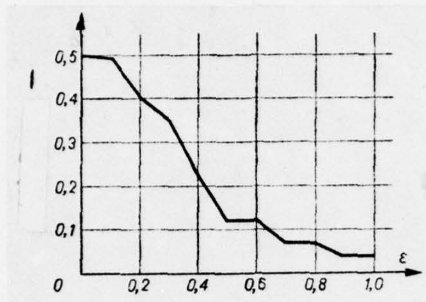


Figure 7. Diagram of the frequency of error as a function of parameter $\mathcal{E}$.

Key: 1-frequency of error.

On the basis of the above equations we can state that the pre-processing layer plays a role in the system adjusting the pattern space to the number given ahead of time for the I inputs of the basic unit of the classifier. The proper choice and orientation of the new system of coordinates has essential influence on both classification mistakes and on the degree of the complexity of the classifier. However, premises do not always exist which permit optimal choice of the number k_0 of linear elements into their parameters.

2.2 Study of Logic Layers

The logic (basic) unit of the classifier is formed by two layers of an adaptive logic network. The first layer of the logic unit transforms k_0-dimensional space into binary space(the k_0L-dimension as high as possible) formed by the apex of the superblock unit. The second layer, called the decision layer, prearranges a decision in conformity with the rule adopted for each apex.

In describing the transformation mechanism performed by the first logic layer, it is convenient to make use of the set of full auxiliary

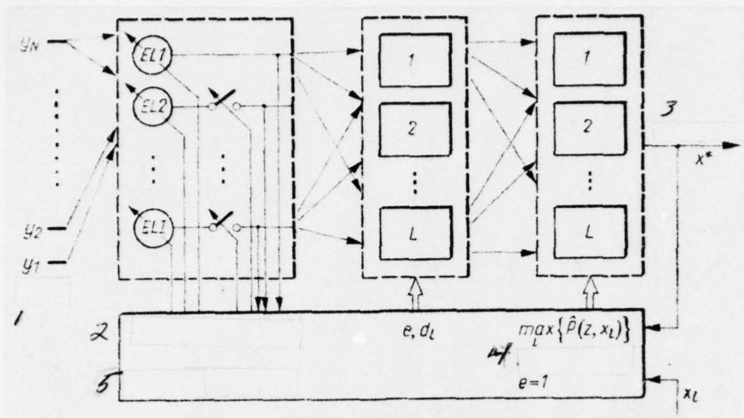


Figure 8. Diagram of cooperation in the cycle studying the classifier with a computer. Key: 1-pattern, 2-reduction in dimensions, 3-decision, 4-for each z e = 1, 5-computer.

---

functions D_{lk} (y' ε). These functions are constructed in consideration of the set studied on the basis of the branch functions defined below

$$\bar{g}_{kj}(y', \varepsilon) = G\{e^{-|(y'-y'[j])\cdot u_k|} - \varepsilon\} \tag{5}$$

where: $y[j]$ is the j observation from the set studied,

u_k is the unit vector along the k coordinate of reduced space,

ε is the extrapolation parameter $\varepsilon \in (0,1]$.

$$G(t) = \begin{cases} 1, \text{where } t \geqslant 0, \\ 0, \text{where } t < 0. \end{cases}$$

$k = 1, 2, \ldots, k_0$.

With the aid of function (5) the process of extrapolating the property of the set studied is fulfilled. It is based on prearrangement of each observation of the environment, the magnitude of which depends on the perameter ε. This approach is a result of the assumption that

[Continued on page 12].

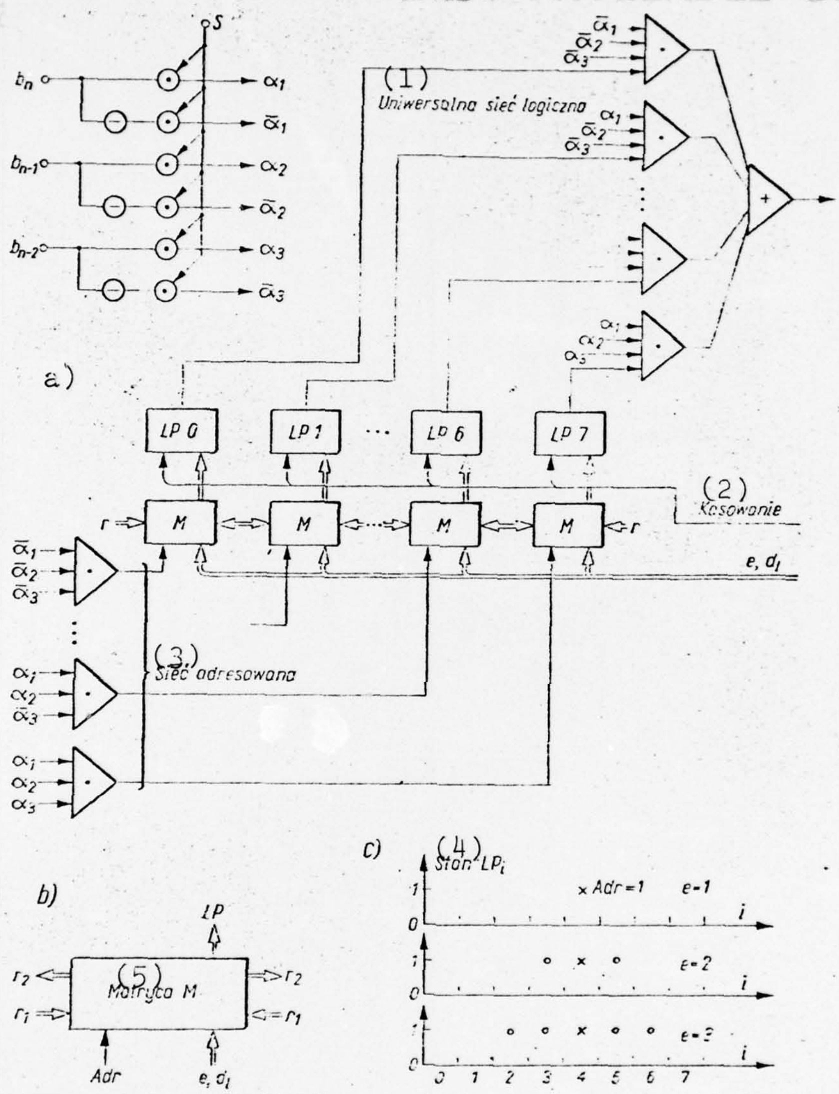


Figure 9. ASL-3 Module: a) Logic diagrams (b_{n-2}, b_{n-1}, input signals, s-module selection signal, LP-memory elements); b) guide matrix, (d_l-block selection signal of class l, e-type of extrapolation, Adr-signal from address network, $r_1 r_2$-signals implementing extrapolation process; c) diagrams illustrating dependence of extrapolation on parameter e(Adr signal switching on LP4 derives from observations in study cycle).
Key: (1) universal logic network; (2) deletion; (3) network addressed, (4) condition LP_i; (5) matrix M.

Table 2. Simplified table of truth of matrix M.

(1) Wejścia				(2) Wyjścia	
e	d_l	Adr	r_1	LP	r_2
0	0,1	0,1	0,1,2	0	0
1	1	1	0	1	0
1	0	1	0	0	0
1	1	0	0	0	0
1	0	0	0	0	0
2	1	1	0	1	1
2	0	1	0	0	0
2	1	0	1	1	0
2	0	0	1	0	0
3	1	1	0	1	2
3	0	1	0	0	0
3	1	0	2	1	1
3	0	0	0	0	0
3	1	0	1	1	0

Key: 1) input; 2) output.

if obserbation Y[j] belongs to $Y(x_l)$, it is very probable that an observation which occurs in the working cycle in the environment y[j] will also belong to this same class.

We obtain the auxiliary function D_{lk} as the logical sum of the proper extrapolation functions

$$D_{lk}(y', \varepsilon) = \mathrm{sgn}\Big[\sum_{Y(x)} d_l \bar{g}_{kj}(y', \varepsilon)\Big] \qquad (6)$$

where $Y(x) = Y(x_1) \vee Y(x_2) \vee \dots Y(x_L),$

$$d_l = \begin{cases} 1, & \text{where } y[j] \in Y(x_l), \\ 0 & \text{in the opposite case,} \end{cases}$$

$l = 1, 2 \dots L.$

Functions (6) can be considered as a certain form of the binary description of the proper of class x_l along the k coordinate.

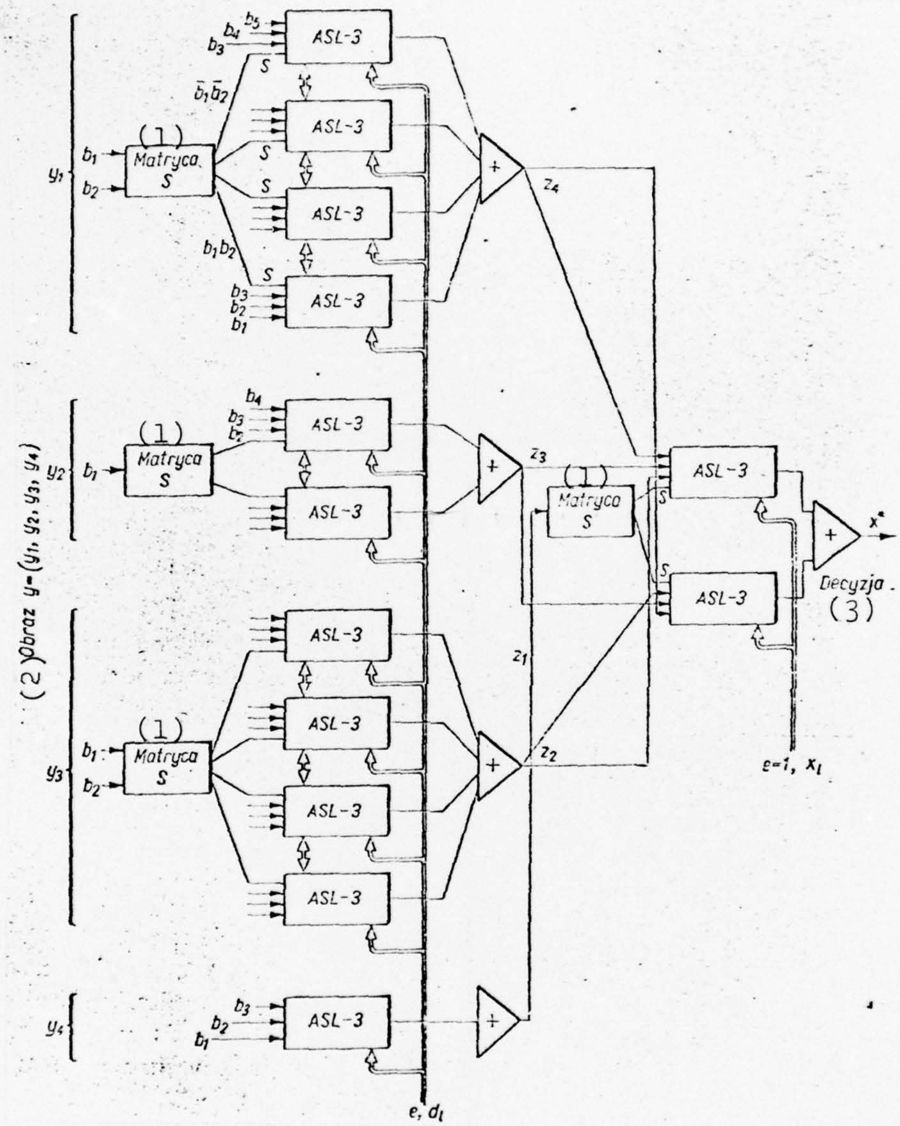


Figure 10. Logic network composed of ASL-3 modules. Key: (1) Matrix S; (2) pattern; (3) decision.

An important moment in the adjustment of the first logic layer is the proper choice of values of parameter ε. The value of this parameter depends on the properties of the set studied.

Figure 2 shows examples of the route of auxiliary functions constructed on the basis of two observations for different values of parameter ε.

In particular for ε=1, it can be seen (fig. 2a) that the auxiliary function assumes a value equal to unity only at definite points corresponding to the values of the k feature of observation z from the study cycle. For $\varepsilon < 1$ the auxiliary function assumes a value equal to unity not only at the definite point, but also in a definite environment of these points (fig. 2b). With further decrease in the values of perameter ε, the environments around the observation will increase and a situation may occur as seen in figure 2c. If $\varepsilon \to 0$, this magnitude of environemnt leads to infinity.

Since the auxiliary functions can assume values 0 or 1, in regards to ~~it~~ them the transformation layer of the logic classifier (fig. 1) transforms the k-dimensional vector y into k_0L-dimensional binary vector z. The dependence between these vectors can be written in the form

$$z = T(y'), \tag{7}$$

where the transformation

$$T(y') = \{D_{11}(y', \varepsilon), \dots D_{lk}(y', \varepsilon), \dots D_{Lk_0}(y', \varepsilon)\} \tag{8}$$

depends on ε, while $D_{lk}(y', \varepsilon)$ means the k auxiliary fuction of the l-class.

It is convienent to explain the activity of the transformation layer by calling on a geometric interpretation. For this purpose let us assume that hyperplanes perpendicular to the proper axes of the set of coordinates are drawn through the places of branches of all auxiliary functions. Then the pattern space will be divided into certain sections, elementary cells. These cells have the property that all vectors y, which belong to a selected elementary cell, will be converted into the same vector z (apices of k_0L-dimensional superblocks). In general the number of elementary cells is greater than the number of vectors z. In this case several elementary cells /488 can be associated with the same apex.

Let us designate by K_z a cell composed of elementary cells to which the same superblock apex number corresponds. Then by the index of cell K_z we can understand the apex number written in the form of a binary number formed from components of vector z. The value of index z in the tenth system will be designated by z'. It can be computed from the formula

$$z' = \sum_{l=1}^{L} \sum_{k=1}^{k_0} D_{lk}(y', \varepsilon) 2^{(l-1)k_0+k-1}. \qquad (9)$$

The study algorithm of the second logic layer makes use of the knowledge of the route of implementation of a variable random Z, together with data about the pertinence of observations to corresponding classes. The histogram $\hat{P}(z, x_l)$, $l = 1, 2, \ldots L$, being an approximation of the distribution probability $P(y \in K_z, x_l)$ is delimited on the basis of such routes. The task which occupies us here is based on empirical evaluation (for all z and classes x_l) of the following probabilities

$$\int_{K_z} p(y, x_l) \mathrm{d}y = P(y \in K_z, x_l) = P(z, x_l), \qquad (10)$$

where $P(y, x_l)$ is the function of probability density of the pair (Y, X),

Y is a variable random constant,

X is a variable random discrete number.

We may accept as an estimate $P(z,x_l)$, the frequencies computed on the basis of sets studied according to the equation

$$\hat{P}(x_l) = \frac{m_l}{\sum_{j=1}^{L} m_j},$$

$$\hat{P}(y \in K_z | x_l) = \frac{m_l[K_z]}{m_l},$$

where: $m[K_z]$ is the number of elements of set $Y(x_l)$ contained in cell K_z,

m_l is the total number of elements in $Y(x_l)$,

$l=1,2,\ldots,L$.

The estimate of the distribution $P(z,x_l)$, $l=1,2,\ldots,L$, represents a histogram which can be taken as a discriminatory function of the classifier, that is $\phi_l(z)=\hat{P}(z,x_l)$. With the aid of this histogram it is possible to define the logic structure of the decision layer and through it to implement a decision rule

$$x^*(y) = x_l, \text{ if } \varphi_l(z) > \varphi_j(z), \quad j = 1, 2, \ldots, L, j \neq l, \tag{11}$$

where the vector z is joined to the vector y by the transformation described earlier.

Rule (11) is a rule approximating the rule of greatest probability, and its quality is characterized by means of frequency of error, which is an estimate of mean probability of error (2).

The activity of the decision layer can also be described by means of auxiliary functions (6). For this purpose the variable y should be replaced by z and it should be assumed that $\mathcal{E}=1$. Treating the variable vector z as a number, we bring our concepts to unidimensional space. In this /1489 case we designate the auxiliary functions for the decision layer by $D_l(z,l)$, $l=1,2,\ldots,L$. In conformity with rule (11), the set of signals represented by the values of the auxiliary functions at the output of the decision layer of the classifier (fig. 1) has the following form

$$D_l(z,l)=\begin{cases}1, & \text{if z corresponds to an observation of the } l \text{ class,}\\ 0 & \text{in the opposite case,}\end{cases}$$

$$l=1,2,\ldots,L. \tag{12}$$

Figure 3 presents a general diagram of the study algorithm along with the classification algorithm.

An important step in finishing the study cycle is an investigation of the quality of classification in two cases:

a) for observations from the study cycle, and

b) for observations from the set doing the examining (known, new, which did not appear in the study cycle).

If, after the study has been made, the classifier gives a bi-passing /1490 decision (neutral), then it is most probable that the cells into which the new observation "falls" is empty. The following may be causes of this:

-the study set does not fulfill the conditions of good representation,

-the choice of the parameter $\mathcal{E}$ is unsuitable (for extrapolation of the properties of the set studied).

We can improve the quality of classification if we repeat the study for a new $\mathcal{E}$ or continue it for observations mentioned in point b). The process of continuing study occurs after the proper conclusion of the study cycle and is based on correction of the frequency histograms (without changing $\mathcal{E}$). In accord with the new histograms changes are made in the structure of the decision layer ~~are made~~. In the case of a change of perameter ε the study embraces two final layers of classifier.

3. Example of Multimodal Classification

The algorithm described will be illustrated by the example of a bidimensional pattern classifier. The study sets for the current example were chosen in such a way that no need to use the preprocessing layer develops. In connection with this we can omit this layer from our consideration and adopt the following assumptions:

1) the classifier is composed of two layers.

2) L=2, that is, two classes of patterns exist, x_1 and x_2.

3) the patterns subject to classification are described by a set of values of two features y_1, y_2, which are realizations of a variable random number within an unknown probability distribution.

4) the observations of each of the classes are extended in such a way that they can be divided into two different sub sets of sets $Y(x_1)$ and $Y(x_2)$. Sets $Y(x_1)$ and $Y(x_2)$ have fifty elements each and are both included in figure 5. The symbol„•" symbolizes observations from class x_1, while the sum symbol „+"observations from class x_2.

5) the decision rule has the form

$$x^*(y) = \begin{cases} x_1 \text{ for } \hat{P}(z, x_1) > \hat{P}(z, x_2), \\ x_2 \text{ for } \hat{P}(z, x_1) < \hat{P}(z, x_2), \\ 0 \text{ for } \hat{P}(z, x_1) = \hat{P}(z, x_2). \end{cases} \qquad (13)$$

(13)

6) the criterion for the quality of the decision rule is the mean frequency of error.

We begin a study of the logic layers of the classifier, the structure of which is shown in fig. 4, by delimiting the set of auxiliary functions for each class. These functions, the values of which are arguments for the decision layer, will be delimited according to equation (6).

Figure 5 gives an example of the division of the pattern space into elementary cells on the basis of auxiliary functions. The diagrams of the appropriate auxiliary functions are set for the parameter $\varepsilon = 0.7$ and are plotted in the figure of both coordinates y_1, y_2.

The binary values, which the auxiliary functions assume, are arranged by the individual elementary cells, in the form of route z. In the example under consideration the following rule was adopted for delimiting /491 the elements of route $= z_1 z_2 z_3 z_4$, where: $z_1 = D_{22}(y, \varepsilon)$, $z_2 = D_{21}(y, \varepsilon)$, $z_3 = D_{12}(y, \varepsilon)$, $z_4 = D_{11}(y, \varepsilon)$. Elementary cells with identical routes z form the cell K_z. An orderly comparison of all the elementary cells from fig.5 (the numbering adopted in the figure is completely arbitrary) is given in table 1.

The study cycle routes of the classifier ends with the construction of a histogram $P(z, x_l)$ $l=1,2$, defined by the cell K_z on the basis of the study set. Fig. 6 shows the histograms formed for this example. On the basis of the histograms it is possible to select a decision layer structure so that the output signal of the classifier has the form:

$$
\begin{aligned}
D_1(z, 1) = 1,\ D_2(z, 1) = 0 \text{ for } \hat{P}(z, x_1) > \hat{P}(z, x_2), \\
D_1(z, 1) = 0,\ D_2(z, 1) = 1 \text{ for } \hat{P}(z, x_1) < \hat{P}(z, x_2), \\
D_1(z, 1) = 0,\ D_2(z, 1) = 0 \text{ for } \hat{P}(z, x_1) = \hat{P}(z, x_2),
\end{aligned}
\qquad (14)
$$

where $D_l(z,1)$, $l=1,2$ designates the auxiliary functions describing the decision layer.

The diagram of frequency of decision error $\overline{P}(\mathcal{E})$, shown in figure 7 leads to some interesting conclusions. It is obvious from this diagram that the frequency of error diminishes when $\mathcal{E}$ tends toward 1. This is caused /492
by the fact that, as $\mathcal{E}$ increases, the auxiliary functions of the transformation layer $D_{lk}(y,\mathcal{E})$, $l,k=1,2$ assume a ~~"comb"~~ "comb-like" form. The dimensions of the cells diminish in connection with this. Consequently the possibility of occurence of cells containing observations of both classes at the same time diminishes. A point corresponds to the border $\mathcal{E}=1$ of each cell which contains at most one observation. This example is useful for deterministic situations when pattern space elements form a final set of points. For statistical situations /493
when observations can form an infinite set of point, it is necessary to extropolate the properties of the classes on the basis of study sets. For this purpose it is necessary to find an adequate $\mathcal{E} < 1$.

It is worthwhile attending to the fact that as it becomes more distant from the points corresponding to the hypothetical modes of the curve of conditional probability density, the approximation of the decision ranges to the cells K_z is worse. At the same time this does not have a great deal of effect of increasing the value of mean probability of error, because the probability of drawing the signal y_1, y_2 from these areas and qualifying it otherwise than would result from the decision rule adopted is relatively slight.

The algorithm described has been used to recognize numbers written by hand. Runs of binary numbers have been arranged for the recognized number in such a way that they correspond to signals from photo elements of the receptor field of dimensions of 5x7 elements. Experimental research with the algorithm has been carried out on the ODRA 1204 computer. Some results of this research are discussed in [4,5].

4. Final Conclusions

Some remarks must be made after this general presentation of the algorithm and of tests using it for classification problems.

One fault in the algorithm lies in the transformation of vector y into binary vector z. The maximal number of components of vector z is the product of the number of dimensions N of pattern space and the number of classes L. In this connection the dimensions of vector z will be large for large N and L. Among other things this can cause prolongation of study time, time for reaching a decision, and difficulties in implementing the algorithm. In regard to the above the reduction of dimensions of the original pattern space, to which the number of components of vector z is immediately related, has great importance. Part II in work [5] is devoted to methods of reducing the dimensions which are part of the problem of reducing data.

The defects mentioned are also partially associated with the fact that implementation of the decision rule is based on histograms. As can be seen from the example presented in subsection 3, it does not pay to reserve space in the computer memory to record histogram coordinates for all combinations of index z, since many of them may not be useful. In this situation it is better to use a method based on alternatives to form histograms for individual classes in the study cycle:

a) increase the actual set of numbers composed of z pairs: the index and the frequency corresponding to it, if there occurs an observation belonging to a cell not yet delimited;

b) correct the frequency, if the observation belongs to one of the cells earlier delimited.

If the number of blocks in the first logic layer is greater than the number L of classes (redundancy), a possibility exists in this situation of assigning more than one block per class, observations of which are particularly unfavorably extended (e.g. with several modes). In this case the redundant blocks have the nature of universal reserve blocks and can be 1494
arbitrarily assigned to chosen classes. Among other things, this is an important virtue in relation to segmental linear classifiers, in which a priori determination of the number of modes alloted to individual classes is required.

A full study cycle of the classifier must include an estimate of two parameters:

a) Δ_0 , on which the number of reduced space dimensions depends, and

b) ε on which the generalizing properties of the classifier depend.

We make the initial choice of parameter ε^2 as a function of the property of observation taking part in the study cycle. If we do not assume the existence of probability distribution, we have at our disposition a series of observations, particularly the study cycle, and it is possible to choose ε close to or equal to unity. In the case where information of a statistical nature is associated with observations, an estimate of the classifier parameter must be carried out from the angle of extrapolation of the properties of the classes on the basis of known observations. This means that it is necessary to choose $\varepsilon < 1$ and to aim at an estimate of ε_{opt}.

Since the decision quality depends on the parameter ε and has an extreme [4,5], it is possible to use empirical methods, such as a final increment algorithm [2,3], to estimate ε_{opt}. In the case of application of methods based on stochastic approximation algorithms, it seems that the limitation of a sign correction, normally based on computation of a derivative, can be replaced by an investigation of the error structure, the relationship between netural and erroneous decisions. In addition it is also possible to use one of the methods of random search to find the minimum $\bar{P}(\mathcal{E})$. In this way the problem of optimizing the decision rule will be reduced to a problem of minimizing the criterion function dependent on estimated perameters.

The generalizing and selective properties of the classifier are associated with the choice of parameter $\mathcal{E}$ and the number of connections between logic layers. In the research we carried out we used all possible connections between layers, even though this is not possible under real conditions. Elaboration of a method of choosing suitable connections for the purpose of reducing their number [3] (or of increasing cell volume) can improve classifier properties as well simplify its structure. Thus the proper method of reducing the dimensions of original pattern space, and then the choice of connections between logic layers, can thoroughly simplify the structure of a classifier and preserve good decision quality.

From the above considerations and from the research carried out [4,5], it is obvious that great possibilities exist for improved pattern classification, both in a wide range of alteration in the properties of observation sets and in the set number of inputs in the logic unit of the classifier. In addition experiments carried out show that it is generally possible, without great difficulty, to obtain correct classification of all observations known from this study cycle. This property is of great importance for deterministic situations where final observation sets occur.

II. POSSIBLITIES OF PRACTICAL IMPLEMENTATION OF THE ALGORITHM 1495

1. Introduction

The construction of a logic classifier, as a completely independent device, is complex and requires further solution of a system guided by the study of individual layers. It seems most reasonable and simple to construct a classifier in the form of hybrid apparatus (fig. 8) composed of a preprocessing layer and adaptive networks (the classifier proper), with a computer (as the guide system in the study cycle). Such factors as the following support this solution:

1) ease in constructing a classifier as equipment cooperating with a computer,

2) possibility of using iterative methods to adjust the preprocessing layer, and

3) the use of binary signals to choose to structure of the logic unit of a classifier.

After the study cycle, the classifier can independently implement the work cycle.

We shall not concerns ourselves with the technical implementation of a preprocessing layer. It is constructed of linear elements, and the difficulties which can occur during implementation are generally known [7,8]. An interesting example of the solution of this element in computer technology, making cooperation with a computer possible, has been worked out and produced in the Bionic Laboratory of the Institute of Applied Cybernetics of PAN [Polish Acamedy of Sciences][6].

2. Implementation of the Basic Unit of a Logic Classifier

2.1 Description of ASL-3 Module

The concept for a technological solution of the basic unit of a classifier given below has been worked out on the basis of the universal ASL-3

module. The ASL-3 module (Fig. 9) is a modification of module W-3 [9] and represents the basic structural unit of a classifier. In contains an adaptive logic network with 3 inputs and 1 output. The universal logic network is formed of /496
8 gates through which all combinations of signals from input b_{n-2}, b_{n-1}, b_n (where n is an index of the last position in the binary expansion of the number) e.g., $y_1 = b_1 b_2 \ldots b_{n-2}\, b_{n-1}\, b_n$) and control signals from flip-flops LP0, LP1, ... LP7 are fed. The outputs from the gates are connected through a logic sum with the output of the module. Just as for module W-3, a special input s serves /497
to expand the area of length of binary numbers introduced into the module. The method of using input s will be illustrated as an example.

LP flip-flops, guided by matrix M and an address network (used exclusively in the study cycle) take part in the control memory of the module. The diagram of matrix M is shown in Fig. 9b in the shape of a black box, and its description is included in Table 2. The class properties are extrapolated with the aid of matrix M on the basis of each observation which occurred in the study cycle. These extrapolation nodes (points on the coordinate axis) are delimited by observations through the medium of signal Adr from the address network in the module. The signal e, a counterpart of parameter $\mathcal{E}$, defines the environment in which property extrapolation takes place. In particular signal e = 1 means simple memorization of observations which occurred in the study cycle. In case $e > 1$, the observations are memorized along with the distinctive environment (Fig. 9). Assignment of observations to a class is determined by signal d_1.

To illustrate the properties discussed, let us look at an example where e = 3. Occurrence of observations at the model input causes matrix M, to which the signals Adr = 1 (from observations) and $d_1 = 1$ are addressed, gives a signal

LP=1 and signal r_2=2 going in two directions. Under the effect of signal r_2^4 in a neighboring matrix (occurring as r_1=2) with ADR=0(d_l=1,e=3), we get LP=1 and r_2=1. The signal r_2=1 in the next matrix causes LP=1, r_2=0 (see last line in table 2). The lack of a signal r_2 means that the process of extrapolation from the observations was completed. In this way the condition of the flip-flops created at the end of the study cycle guarantees implementation of a definite auxiliary function by the logic network. In the work cycle, the lack of signal e(e=0) protects flip-flops LPi,i=0,1,...,7 from changes in condition (see first line in table 2).

2.2. Construction of Logic Network 1498

With suitable connection of modules we get a complex network in which the number of binary inputs for each component of the vector of observation can be greater than 3. Fig. 10 shows an example of a bilayer logic network constructed of ASL-3 modules. This network can serve as a binary classifier for patterns $y=(y_1,y_2,y_3,y_4)$. The values of signals corresponding to individual pattern components are binary numbers and have the following shape

1499

$$\begin{aligned} y_1 &= b_1b_2b_3b_4b_5, \\ y_2 &= b_1b_2b_3b_4, \\ y_3 &= b_1b_2b_3b_4b_5, \\ y_4 &= b_1b_2b_3, \end{aligned} \qquad (1)$$

where: b_j=0 or 1,j=1,2,...,5.

The S matrices in the network presented work as commutators. The module, in which a logic unit is found at an additional input, will be chosen as a function of the condition of signals b_1 and b_2 at their input (we shall explain this for component y_1). This unit opens a route for further signals b_3, b_4, b_5 of component y_1. In this way the breakdown of the permissable set of values of component y_1 into four subsets, with the aid of matrix S, corresponds to the choice of module

$$\begin{array}{cccc} 00\,000 & 01\,000 & 10\,000 & 11\,000 \\ 00\,001 & 01\,001 & 10\,001 & 11\,001 \\ \cdots & \cdots & \cdots & \cdots \\ 00\,111 & 01\,111 & 10\,111 & 11\,111 \end{array} \qquad (2)$$

Each subset (2) is a set of values of arguments for a suitable module.

A second layer of the logic unit of a (decision) classifier can be composed of exactly the same ASL-3 as the first layer. The input signal of the second layer is a binary vector z formed of output signals from the first layer (see figure 10, $z=(z_1, z_2, z_3, z_4)$. The basic difference between the first and second layers is the following. For the first layer every observation y can be treated as a point in the K_0-dimensional space of features, the coordinates of which assume discrete values from a certain section. In the case of the second layer the vector of signal z can be treated as a point in K_0L-dimensional binary space (apex of a unit Superblock). In addition the output signal of the second layer is simultaneously the output signal for the entire network.

The structure of the second layer will be chosen at e=1, in conformity with rule (1)(part I), and the quality of the decisions organized by the superblock apices is determined with the aid of criterion (2) (part I).

3. Final Considerations

An important feature in the technological solution of the basic unit of a classifier is the possibility of achieving it in the form of a logic network with a modular structure. This means that both network layers can be composed and expanded with the same ASL-3 modules.

The construction of the ASL-3 module as a computer system does not present any serious difficulties. The two-phase work of computer systems is associated with rather slight requirements in element tolerance and causes them to be suitable for technological realization, particularly in the form of self-contained circuits. In addition the possibility of using self-contained circuits for this purpose brings about other advantages, such as great reliability, small production costs, miniaturization, great rapidity of computer system activity, etc.

One shortcoming in the solution presented for the basic unit of 1500
a classifier is the expanded structure of the ASL-3 module (effect of universality). Redundant construction of the decision layer occurs in connection with this, since in this case the modules are used only for e=1. In addition the lower flexibility (than for a programmed version) in choosing the extrapolation perameter e can cause greater classification errors in certain situations.

In the general case the basic unit can work as an adaptive logic network, and particularly as a pattern classifier.

Working out the linear (threshold) element, in contradistinction to logic systems, can present some difficulties. These difficulties occur mainly from the need to use non-typical solutions in the construction of altered weights of linear elements [7,8]. The use of computer technology for this purpose turns out to be very advantageous from many points of view [6]. In general classification systems in which the linear element is one of the main components (e.g., segmental linear classifiers, threshold layer classifiers) are more expedient in program implementation than in systems implementation.

Literature

[1] **Mattson R., Dammann J.**: *A Technique for Determining and Coding Subclasses in Pattern Recognition Problems.* I.B.M. J., July, 1965.

2 Seidler, J, "Optimizations of adaptive information systems", PWN, Warsaw, 1971.

3 Rutkowski, D. "Random search for place of minimum functioning of many variables". Joint work "Optimal information retrieval". PAN Material Conference, Ossolineum Wroclaw, 1968,

4 Malina W., W. Sobczak, "Experiemental results with an algorithm for number recognition", Conference Material, "Methods of direct insertion and retrieval of text information", 17–19.X.73

5 Malina W. "Doctoral work", 1974

6 Gawronski R., "Layer networks composed of pseudoneural elements applied to preprocessing of data". Work of the Institute of Applied Cybernetics of PAN, Warsaw, 1972,

7 Kulikowski, J. "Cybernetics recognition systems", PWN, Warsaw, 1972

[8] **Smith J., Harbourt C.**: *An Adaptive Threshold Logic Gate Using Capacitive Analog Weights.* IEEE Trans. on C., Jun. 1968.

9 Malina, W. Adaptive converter with a limited number of outputs to machine Ø., 1972,

Footnotes

1 Detailed description of the present algorithm and an example of measuring set coverage can be found in [5].

2 In general $\mathcal{E}$ can be a vector of component ($\mathcal{E}_1$, $\mathcal{E}_2$, ..., $\mathcal{E}_{k0}$).

3 An example of the method of choosing connections based on the algorithm of pre-processing layer adaptation is discussed in point 2.1. given in [5].

4 Attention should be called to the fact that the output signal r_2 of one matrix is the input signal r_1 of a neighboring matrix.

A NON-PARAMETRIC ALGORITHM FOR PATTERN CLASSIFICATION

Summary

The classification algorithm described in part I can be used in the case of incomplete initial information. From the functional point of view in the classifier there are distinguished: the preprocessing layer, the transformation layer and the decision layer. During learning a rule was applied according to which particular classifier layers are subsequently adjusted.

The main feature of the algorithm concerns the partitioning of pattern space into cells. Then for each class a histogram is deviced being an estimate of an unknown distribution of the probability of appearing of the pattern from particular cells. The way of building cells and histograms takes into account the form of initial information. The details concerning the algorithm are explained on an example.

In the part II there was given a concept of practical implementation of the classifiers basic unit in the form of two layers of adaptive logical networks. Each layer of logical network can be assembled of modelas ASL-3. The module ASL-3 (Fig. 9) is the basic construction unit of the classifier. It is an adaptive logical network with three inputs and one output. In the case of building the classifier there exist a possibility of connecting it simply for working with a computer.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER FTD-ID(RS)I-1209-76	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A NON-PARAMETRIC ALGORITHM FOR PATTERN CLASSIFICATION		5. TYPE OF REPORT & PERIOD COVERED Translation
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Witold Malina		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Foreign Technology Division Air Force Systems Command U. S. Air Force		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 1975
		13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

09; 12

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DISTRIBUTION LIST

DISTRIBUTION DIRECT TO RECIPIENT

ORGANIZATION		MICROFICHE
A205	DMATC	1
A210	DMAAC	2
B344	DIA/DS-4C	8
C043	USAMIIA	1
C509	BALLISTIC RES LABS	1
C510	AIR MOBILITY R&D LAB/FIO	1
C513	PICATINNY ARSENAL	1
C535	AVIATION SYS COMD	1
C557	USAIIC	1
C591	FSTC	5
C619	MIA REDSTONE	1
D008	NISC	1
H300	USAICE (USAREUR)	1
P005	ERDA	2
P055	CIA/CRS/ADD/SD	1
NAVORDSTA (50L)		1
NAVWPNSCEN (Code 121)		1
NASA/KSI		1
544 IES/RDPO		1
AFIT/LD		1

ORGANIZATION		MICROFICHE
E053	AF/INAKA	1
E017	AF/RDQLR-W	1
E404	AEDC	1
E408	AFWL	1
E410	ADTC	1
E413	ESD	2
	FTD	
	CCN	1
	ETID	3
	NIA/PHS	1
	NICD	5