INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - I

H4.SMR/303 - 3

WORKSHOP
GLOBAL GEOPHYSICAL INFORMATICS WITH APPLICATIONS TO RESEARCH IN EARTHQUAKE PREDICTIONS AND REDUCTION OF SEISMIC RISK

(15 November - 16 December 1988)

PATTERN RECOGNITION. GENERAL DESCRIPTION

V.I. KEILIS-BOROK
&
A. A. SOLOVIEV

Institute of Physics of the Earth
Academy of Sciences of the U.S.S.R.
Bolshaya Gruzinskaya 10
123 242 Moscow
U.S.S.R.

EXAMPLES OF PATTERN RECOGNITION PROBLEMS

1. Recognition of earthquake-prone areas.

To divide the selected structures of a region into two classes:

-- structures where earthquakes with $M \geqslant M_o$ may occur;

-- structures where earthquakes with $M \geqslant M_o$ may not occur.

2. Recognition of strata filled with oil.

To divide the strata encountered by a borehole into two classes:

-- strata which contain oil;

-- strata which contain water.

3. Medical diagnostics

To divide examined people into two classes:

-- people who have a specific disease;

-- people who do not have it.

GENERAL FORMULATION OF THE PATTERN RECOGNITION PROBLEM

Generalizing the three above examples one may formulate the problem of pattern recognition abstractly as follows:

The set $W = \{ w^i \}$ is considered, where objects $w^i = (w^i_1, w^i_2, \ldots , w^i_m)$, $i=1,2,\ldots$ are vectors with real (integer, binary) components.

The problem is to divide the set W into two or more subsets which differ in certain feature or according to clustering themselves.

Two kinds of pattern recognition problems and methods:

-- classification without learning;

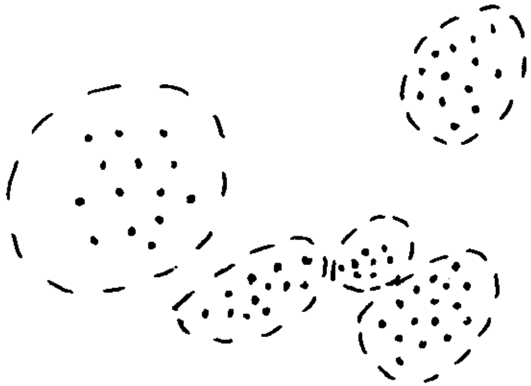
-- classification with learning.

CLASSIFICATION WITHOUT LEARNING
(CLUSTER ANALYSIS)

The set W is divided into groups (clusters) on the base of some measure in the m-dimennsional space $w_1, w_2, \dots, w_m$.

Denote $\rho(w, v)$ a distance between two m-dimensional vectors $w = (w_1, w_2, \dots, w_m)$ and $v = (v_1, v_2, \dots, v_m)$.

To estimate quality of classification the special function is introduced. The best classification gives the extremum of this function.



Examples of functions

W is a finite set.

$$J_1 = \frac{(K-1)\sum_{k=1}^{K} \rho_k}{2\sum_{k=1}^{K-1}\sum_{j=k+1}^{K} \rho_{kj}} \longrightarrow \min,$$

$$J_2 = \frac{1}{K}\left[\sum_{k=1}^{K}\rho_k - \frac{2}{K-1}\sum_{k=1}^{K-1}\sum_{j=k+1}^{K}\rho_{kj}\right] \longrightarrow \min.$$

Here K is the number of groups

$$\rho_k = \frac{2}{m_k(m_k-1)}\sum_{i=1}^{m_k-1}\sum_{j=i+1}^{m_k}\rho(w^i, w^j),$$

$$\rho_{kj} = \frac{1}{m_k m_j}\sum_{i=1}^{m_k}\sum_{s=1}^{m_j}\rho(w^i, v^s),$$

m_k, m_j are the numbers of objects in the k-th and j-th groups respectively,

$w^1, w^2, \dots, w^{m_k}$ are the objects of the k-th group,

$v^1, v^2, \dots, v^{m_j}$ are the objects of the j-th group.

After the groups are formed the next problem can be formulated: to find common feature of objects belonged to the same group.

As a rule the set W is divided into two classes, say D and N .

The apriori examples of objects of each class are given. They are called the learning set:

$$W_o \subset W,$$

$$W_o = D_o \sqcup N_o,$$

D_o is the learning set of objects belonging to class D,

N_o is the learning set of objects belonging to class N.

The result of the pattern recognition is twofold:

-- The rule of recognition; it allows to recognise which class an object belongs to knowing the vector w^i describing this object.

-- The actual division of objects into separate classes according to this rule:

$$W = D \sqcup N .$$

In some cases there are objects with undifined classification, so

$$W = (D \sqcup N) \sqcup U .$$

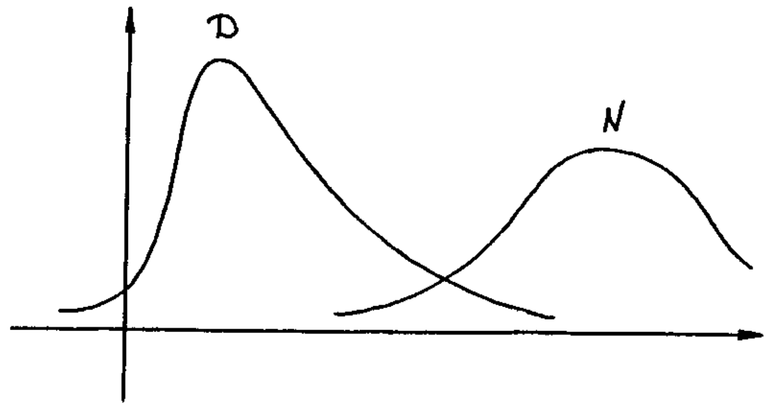
Analysis of the obtained rule of recognition may give information about the connection between the feature which differs the classes D and N on one hand and description of objects (components of vectors w^i) on another.

For example, analysis of the rule for recognition of earthquake-prone areas gave an insight of the role of fluids in the origin of the earthquakes.

EXAMPLES OF ALGORITHMS

1. STATISTICAL

It is assumed that distribution laws are different for vectors from classes D and N.



Samples D_o and N_o are used to define the parameters of these laws.

The recognition rule is based on calculation of an estimation of conditional probabilities for each object w^i to belong to class D (P^i_D) and N (P^i_N) and a clasification of the object according to these probabilities:

$$w^i \in D \quad \text{if} \quad P^i_D - P^i_N \geq \varepsilon ,$$

$$w^i \in N \quad \text{if} \quad P^i_D - P^i_N < -\varepsilon ,$$

$$w^i \in U \quad \text{if} \quad | P^i_D - P^i_N | < \varepsilon ,$$

where $\varepsilon \geqslant 0$ is a given constant.

An example of a statistical algorithm is Bayes algorithm.

BAYES ALGORITHM

According to Bayes formula

$$P(w=w^{i}|w\in D)\ P(w\in D)\ =\ P(w\in D|w=w^{i})\ P(w=w^{i})$$

and

$$P_{D}^{i}\ =\ P(w\in D|w=w^{i})\ =\ \frac{P(w=w^{i}|w\in D)\ P(w\in D)}{P(w=w^{i})}\ .$$

Similarly

$$P_{N}^{i}\ =\ P(w\in N|w=w^{i})\ =\ \frac{P(w=w^{i}|w\in N)\ P(w\in N)}{P(w=w^{i})}\ .$$

Estimations of probabilities in the right side of these relations are given by following approximate formulae in which the samples D_o and N_o are used:

$$P(w=w^{i}|w\in D)\approx P(w=w^{i}|w\in D_o)\ ,$$

$$P(w=w^{i}|w\in N)\approx P(w=w^{i}|w\in N_o)\ ,$$

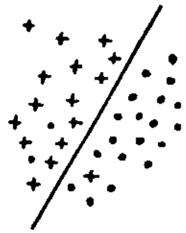
$$P(w=w^{i})\approx P(w=w^{i}|w\in D_o)\ P(w\in D)\ +\ P(w=w^{i}|w\in N_o)\ P(w\in N)\ .$$

Probability $P(w\in D)$ is a parameter of the algorithm and has to be given, $P(w\in N)\ =\ 1\ -\ P(w\in D)$.

NOTE: The sign of the difference $P_{D}^{i}\ -\ P_{N}^{i}$ does not depend of value $P(w=w^{i})$.

2. GEOMETRICAL

In these algorithms surfaces in the space $w_1,\ w_2,\ \ldots,\ w_m$ are constructed to separate classes D and N .



An example of a geometrical algorithm is the algorithm Hyperplane.

ALGORITHM HYPERPLANE

In the space $w_1,\ w_2,\ \ldots,\ w_m$ the hyperplane

$$P(w)\ =\ a_1 w_1\ +\ a_2 w_2\ +\ \ldots\ +\ a_m w_m\ +\ a_o$$

is constracted.

This hyperplane has to separate the sets D_o and N_o by the best way. It means that some function has to have extremum value.

The example of the function is

$$J(a_o, a_1, \ldots, a_m)\ =\ \sum_{i=1}^{n_1} P(w^{i})\ -\ \sum_{i=1}^{n_2} P(v^{i})\ \longrightarrow\ \max\ ,$$

where $w^1,\ w^2,\ \ldots\ ,\ w^{n_1}$ are objects of D_o ,

$v^1,\ v^2,\ \ldots\ ,\ v^{n_2}$ are objects of N_o .

The recognition rule is the following:

$w^{i}\in D$ if $P(w^{i})\ \geq \varepsilon$,

$w^{i}\in N$ if $P(w^{i})\ <\ -\varepsilon$,

$w^{i}\in U$ if $|\ P(w^{i})\ |\ <\ \varepsilon$,

where $\varepsilon \geq 0$ is a given constant.

3. LOGICAL

In these algorithms characteristic traits of classes D and N are searched using the sets D_o and N_o. Traits are boolean functions on $w_1, w_2, \dots, w_m$. The object w^i has the trait if the value of the corresponding function calculated for components of the vector w^i is true and does not have the trait if it is false. A trait is a characteristic trait of the class D if objects of the set D_o have this trait more often than objects of the set N_o. A trait is a characteristic trait of the class N if objects of the set N_o have this trait more often than objects of the set D_o.

Using the seached characteristic traits the recognition rule is formulated by following way

$$w^i \in D \quad \text{if} \quad P_D^i - P_N^i \geq \Delta + \varepsilon,$$

$$w^i \in N \quad \text{if} \quad P_D^i - P_N^i < \Delta - \varepsilon,$$

$$w^i \in U \quad \text{if} \quad | P_D^i - P_N^i - \Delta | < \varepsilon,$$

where P_D^i and P_N^i are the numbers of characteristic traits of classes D and N which the object w^i has, Δ and $\varepsilon \geq 0$ are given constants.

Logical algorithms are useful to apply in cases then the numbers of objects in sets D_o and N_o are small.

As a rule logical algorithms are applied to vectors with binary components. An example of logical algorithm is the algorithm CORA. It is applied to geophysical problems in particular to the problem of long-term prediction of earthquake and will be described in details in the lecture "Pattern recognition in earthquake prediction".

DISCRETIZATION AND CODING

As it was mentioned above some pattern recognition algorithms can be applied only to vectors with binary components. In the case when the set W initially consists of vectors with real components the discretization and coding of components are necessary.

After discretization the data become robust. For example if a range of some component is divided into three parts only three gradations for this component ("small", "medium", "large") are used after the discretization instead of its exact value. Don't regret the loss of information. This makes results of recognition stable to variations of data.

TESTS FOR ESTIMATION OF RELIABILITY OF RESULTS

These tests are necessary to be sure in the obtained results. It is especially important in the case of small samples D_o and N_o. The tests illustrate - how reliable are the results of the pattern recognition. However they do not provide a proof in the strict statistical sense if the learning material is small.

The examples of some tests are listed below.

1. To save the part of objects from W_o for recognition only, not using it in learning.
2. To check the conditions: $D_o \subset D$, $N_o \subset N$.

NOTE: Sometimes this conditions are not valid becaus the sets D_o and N_o are not "clear" enough. For example in the case of recognition of earthquake-prone areas objects of D_o are structures where epicenters of earthquakes with $M \geqslant M_o$ are known and objects of N_o are structures where epicenters of such earthquakes are not known. Objects of N_o may belong to the class D, because in some areas earthquakes with $M \geqslant M_o$ may be possible, though yet unknown. Objects of D_o may belong to the class N due to the errors in catalog (in epicenters and/or magnitude).

NUMERICAL TESTS. These tests include some variation of the objects, used components of vectors, numerical parameters etc. The test is positive if the results of recognition are stable to these variations.

3. Elimination of objects from D_o and N_o - one at a time. Formal criteria of stability - small value of the ratio $\frac{m_D}{|D_o|}$ or $\frac{m_D + m_N}{|D_o| + |N_o|}$. Here m_D and m_N show how many objects of D_o and N_o respectively change classification after they were eliminated from learning.
4. Learning on the subsets of the obtained sets D and N.
5. Change the set of used components. In particular elimination of each used component in turn.

Since the danger of selfdeception is not completely eliminated by these tests the design and implementation of new tests should be pursued.

NOTE: The lecture "Pattern recognition in earthquake prediction", the exercises and the user's guide for the package of pattern recognition programs contain the additional information about the subject of this lecture.