INFORMATICA, 1991, Vol.2, No.1, 077–099

DISCRIMINANT ANALYSIS OF RANDOM PROCESSES

Algirdas–Mykolas MONTVILAS, Ryshard NORKOWICH, Edward SHPILEWSKI

Institute of Mathematics and Informatics,
Lithuanian Academy of Sciences,
232600 Vilnius, Akademijos St.4, Lithuania

Abstract. The problem multialternative recognition of non-stationary processes on the basis of dynamic models is investigated in the paper. The algorithms of pointwise and group classifications are compared. Clustering algorithms based on nonlinear mapping of the segments of random processes onto the plain are used to construct the classifiers.

Key words: recognition of non-stationary sequences, current and group classification, clustering.

Introduction. Receiving, information processing and development of quided effects is basic in the work of complex objects control systems, in particular of automatic and automated management and control systems. During information processing a control system has to solve such problems as detection, coordinate and parameter estimation, smoothing, prediction, recognition, moment determination of change in properties and other. The problem of noise, vibrational, radio and hydrolocational signal recognition arises in technical and medical diagnostics, during the quality control of systems according to the dynamic data under the control of techno-

logical processes, while recognizing moving dynamic objects and in a number of other problems. The recognition of stationary random processes based on a description of classes by dynamic models in the shape of differential and difference stochastic equations was firstly considered in the papers of Shpilewski (1971), Petrov and Shpilewski (1974). Further in connection with the development of digital computers linear autoregression models AR gained a wide dissemination in practice (Shpilewski, 1980). The proposed approach to the recognition of stationary random processes appeared to be extremely fruitful. It allowed to obtain the constructive methods of recognition of random processes with a continuous and discrete time, to work out dynamic classification algorithms in the recognition of sequences in a current time, in the recognition under noise circumstances and partial observation. In the paper by Krutowski and Shpilewski (1975) on the basis of the approach proposed the problem of a sequential multialternative recognition, detection and moment estimation of a change in properties of the combined process has been solved.

Training algorithms in the recognition under the conditions of an apriori indetermination have been developed. The dependence of a recognition error on the length of training and recognitable stationary sequences has been investigated (Shpilewski, 1986).

An urgent problem is the recognition of non-stationary random processes. One of the approaches consists of the use of a piecewise-stationary approximation (Atamukas and Shpilewski, 1979) of non-stationary processes by the stationary dynamic models, in particular, by AR models. The approach under consideration is reduced to a separation of the non-stationary process to subclasses of stationary segments and a description of each subclass by a stationary model. This approach is also used for the recognition of non-stationary processes, presenting by itself a mixture of the finite number

of stationary processes. Cluster analysis is used for the construction of subclasses.

The method based on the construction of subclasses allows as to obtain a constructive description of subclasses of non-stationary processes and construct decision rules for the classifiers. Under this approach general number of classes essentially increases and a necessity arises for the recognition of segments of the sequence differing in length. The use of recurrent equations in the classifiers for sufficient statistics which allow to obtain a pointwise processing of the sequences and realize the recognizing device in the shape of a multichanuel conveyer operator is non-advisable, i.e., computer time, necessary for the recognition of a sequence considerably increases. A necessity arises to develop group classification algorithms, allowing to reduce processing and decision making time.

2. Statement of the problem. Let in the measurable space $(\Omega, \mathcal{F})$ a random process Y_t, $t \geqslant 0$ with a continuous time $Y_t = \{y_r, 0 \leqslant r \leqslant t\}$ or with a discrete time $Y_n = \{y_1, y_2, \ldots, y_n\}$ be defined. Let $\{\mathcal{F}_t\}$, $t \geqslant 0$ be a non-decreasing set of σ-algebras $\mathcal{F}_t = \sigma\{y_s, s \leqslant t\}$ and $\mathcal{F} = \sigma\{U\mathcal{F}_t\}$, with respect to which y_s, $s \leqslant t$ is measurable. With respect to the observed process y_t M alternative hypotheses were introduced

$$H = \{h_1, h_2, \ldots, h_M\}, \tag{1}$$

making up the whole group of events, i.e., if p_m is an apriori probability of the hypothesis h_M, then

$$p_m > 0, \quad \sum_{m=1}^{M} p_m = 1. \tag{2}$$

M probability measures $P^{(m)}$, $m = \overline{1, M}$ are connected with the hypotheses, defined in $(\Omega, \mathcal{F})$ and the measure

$P = \sum p_m P^{(m)}$. Let $P_t, P_t^{(m)}$ be a contraction of the measures P and $P_t^{(m)}$ on σ-algebras $\mathcal{F}_t$. Each random process Y_t of class m induces the measures $\mu_m(A_t)$ in the space of realizations

$$\mu_m(A_t) = P^{(m)}\{w : \ Y_t \in A_t\}.$$

Everywhere further the dependence of the measure $P^{(m)}$ on index m may be both functional and parametric. In the latter case probability characteristics depend on the vector of the parameters Θ_m.

The problem of multialternative classification of the observed realization

$$y_t^0 = \{y_\tau, \ 0 \leqslant \tau \leqslant t\} \tag{3}$$

in the case of a continuous time or

$$y_n^1 = \{y_1, y_2, \ldots, y_n\} \tag{4}$$

in the case of a discrete time consists in a partitioning of the space of y_n realizations (respectively Ω) to non-overlapping sets A_m (or the space Ω to w_m), $m = 1, 2, \ldots, M$ from the minimum condition of probability error

$$\min P_{er}(t) = \min \sum_{m=1}^{M} P^{(m)}\{w : y_t^0 \bar{\in} A_m(t), \ h_m\}$$

for the decision rule: the observed realization y_t^0 relates to class h_k if

$$y_t^0 \ \in \ A_k.$$

3. General solution. The construction of abstract sets $A_k(t)$ in the space of realizations or w_k in Ω may be obtained using decision rules

$$\delta(h, y_t^0)$$

and the theory of statistical solutions (Wald, 1946).

The randomized decision rule $\delta(h, y_t^0)$ for each y_t^0 defines the probability $\delta(h_j, y_t^0)$ in each of the possible classes h_j

$$\sum_{i=1}^{M} \delta(h_j, y_t^0) = 1.$$

For the loss function $l_{ij} = l(h_i, h_j)$ mean losses equal to

$$\sum_{j=1}^{M} l(h_i, h_j)\, \delta(h_j, y_t^0),$$

and the risk $R_\delta(i)$ is determined as a mean value

$$R_\delta(i) = \int_\Omega \sum_{j=1}^{M} l(h_i, h_j)\, \delta(h_j, y_t^0)\, P_i(dw). \tag{5}$$

The function $R_\delta(t)$ defines M risk values. In this case there arises a problem to find out decision rule (5), minimizing the maximal risk value. We come to the minimax criteria of classification. Under Bayesian approach when all apriori probabilities (2) of classes $h_1, h_2, \ldots, h_M$, are known, we calculate an average risk

$$R_\delta = \sum_{i=1}^{M} p_i \int_\Omega \sum_{j=1}^{M} l(h_i, h_j)\, \delta(h_j, y_t^0)\, P_i(dw). \tag{6}$$

An optimal decision rule is defined from the average risk R_δ minimum condition.

It is abvious that the average risk will be minimal, if decision rules are selected from the minimum condition of the aposteriori risk for each realization of observed process (3).

$$R_\delta(y_t^0) = E[l(h_i, \gamma(y_t^0))/y_t^0]. \tag{7}$$

Introduce an aposteriori probability of the hypothesis h_i by the relation

$$p_i(t) = \frac{dP_i}{dP}(w)p_i.$$

Then the aposteriori risk is equal to

$$R_\gamma(y_t^0) = \sum_{i=1}^{M} l(h_i, \gamma(y_t^0))\, p_i(t). \tag{8}$$

Hence, it follows that the aposteriori risk will be minimal, if the observed realization y_t^0 is refered to such a class h_k, for which

$$h_k : \sum_{i=1}^{M} l_{ik}p_i(t) \leqslant \sum_{i=1}^{M} l_{ij}p_i(t), \quad j = \overline{1, M}. \tag{9}$$

The average risk criterion of minimum is a sufficiently general criterion. The criterion of ideal observation, the criterion of weighted combination and the criterion of Neyman–Pearson are partial cases, e.g. in the detection problem (recognition of two classes). These criteria are obtained from the average risk general criterion of minimum (6) for different loss functions. For an ordinary loss function of the shape

$$l_{ij} = \begin{cases} 1, & i \neq j, \\ 0, & i = j, \end{cases}$$

the average risk criterion minimizes the complete probability of misclassification. The decision rule will take the shape

$$h_k : p_k(t) = \max_i p_i(t). \tag{10}$$

The considered general mathematical statement of a process recognition problem and the standart solution within the frames of the theory of statistical solutions extremely simplify the problem and conceal those difficulties which arise while solving practicular problems. In order to build the constructive methods for the recognition of real processes two methods have an essential value:

– in what shape the measures P_i are set for the description of real processes;

– what degree of indeterminacy is in setting of these measures.

4. Mathematical models of random processes. The methods of recognition of random signals based on the use of stochastic dynamic models for the description of processes of the classes are considered in the given paper. In the case of a continuous time the processes of class h_m are described by the stochastic differential equations

$$h_m : dy_t = A_m(w, y, t)\, dt + B(y, t)\, dW_t, \; m = \overline{1, M}, \quad (11)$$

where $A_m(w, y, t)$ is an l-dimensional random vector function of time t, observations $y = y_t^0$, random event w, $B(y, t)$ is the matrix of dimensionality functions $[l \times l]$, W_t is an l-dimensional standart Wiener process.

In the case of a discrete time the processes of class h_m are described by the recurrent stochastic equations

$$h_m : y_n = A_m(y_{n-1}^{n-p}, n) + B_m(y_{n-1}^{n-p}, n) V_i(n), \; m = \overline{1, M}, \quad (12)$$

where $A_m(y_{n-1}^{n-p}, n)$ is the l-dimensional random function, $B_m\ (y_{n-1}^{n-p}, n)$ is the matrix of dimensionality functions $[l \times l]$, $V_i(n)$ is a sequence of independent normally distributed random variables.

In particular, for the description of classes of stationary processes, linear autoregression models are used in the shape

$$h_m : y_n = a_0^{(m)} + \sum_{i=1}^{p} a_r^{(m)} y_{n-r} + b_m V_n^{(m)}, \ m = \overline{1, M}. \quad (13)$$

Parameters of the dynamic models are defined in the learning regime by the realizations of random processes of known classes. For this purpose the methods of the least squares maximal likelihood, Bayes or Yule–Walker equations are used.

5. Classification of processes with a continuous time. The main equations for the statistics used for the recognition of classes of the processes, described by models (11), we shall get on the basis of the following theorem.

Theorem. *Let an observed process y_t be the Ito process and for each hypothesis h_m be defined by stochastic differential equation (11), the coefficients of which satisfy usual limitations, accepted in the optimal nonlinear filtration theory* (Liptser, Shiryaev, 1974, Theorem 8.1). *Then the aposteriori probability of class $p_m(t)$ satisfies the differential equation*

$$dp_m(t) = p_m(t)[\bar{A}_m(t) - \bar{A}(t)]^* (BB^*)^{-1}[dy_t - \bar{A}(t)dt] \quad (14)$$

with the initial conditions

$$p_m(0) = p_m,$$

where $\bar{A}_i(t)$, $\bar{A}(t)$ are determined by the relation

$$\bar{A}_m(t) = E[A_m(w, y, t)/y_t^0, h_i],$$

$$\bar{A}(t) = \sum_{i=1}^{M} A_i(t) p_i(t).$$

We get the proof of the theorem from a general equation of nonlinear filtration. We shall get the equations for the relation of aposteriori probabilities $\nu_{ij}(t)$ and the likelihood relation $\Lambda_{ij}(t) = \frac{p_i}{p_j}\nu_{ij}(t)$ using Ito change of variables formula in the shape

$$d\nu_{ij}(t) = \nu_{ij}[\bar{A}_i(t) - \bar{A}_j(t)]^*(BB^*)^{-1}[dy_t - \bar{A}_j(t)dt] \quad (15)$$

with the initial conditions

$$\nu_{ij}(0) = p_i/p_j.$$

It is possible to obtain in a similar way the differential equations for a likelihood relation logarithm, for the likelihood function and other statistics, used for process recognition.

The sufficient statistics $p_m(t)$, $\nu_{ij}(t)$ are determined as a result of solution of differential equations (14), (15), in this connection we face an essential difficulty implied in the calculation of an aposteriori mathematical expectation $\bar{A}_m(t)$. In partial cases it is possible to write out evident expressions for the calculation of $\bar{A}_m$. If the random function $A_m(w, y, t)$ in (11) is defined as a function of an unobserved component x_t in the shape $A_m(w, y, t) = A_m(x_t, y_t, t)$, then the problem is reduced to the evaluation of the aposteriori distribution x_t. It is possible to get a constructive solution when the unobserved component x_t enters linearly (11) and is reduced to the calculation of the aposteriori mathematical expectation, i.e., to the filtration problem.

6. Pointwise classification in the recognition of Markov sequences. Consider M classes of Markov processes with a transient probability density of observations y_n of the shape

$$h_m : f_m(y_n/y_{n-1}^{n-p}, n), \quad m = \overline{1, M}, \quad (16)$$

and the probability distribution density of the initial vector y_p^1

$$h_m : f_m(y_p^1), \quad m = \overline{1, M}. \tag{17}$$

We shall get the solution of Markov sequences recognition problem (16), (17) on the basis of recurrent equations for the aposteriori probabilities of classes in the shape

$$h_m : p_m(n) = \frac{f_m(y_n/y_{n-1}^{n-p})p_m(n-1)}{\sum\limits_{i=1}^{M} f_i(y_n/y_{n-1}^{n-p})p_i(n-1)}, \quad m = \overline{1, M}, \tag{18}$$

with the initial conditions

$$p_m(p) = \frac{p_m f_m(y_p^1)}{f(y_p^1)}.$$

For the likelihood function $L_i(n)$ the recurrent equations follow from (18)

$$h_m : L_m(n) = f_m(y_n/y_{n-1}^{n-p})L_m(n-1), \quad m = \overline{1, M}, \tag{19}$$

with the initial conditions

$$L_m(p) = f_m(y_p^1).$$

Obviously processes (12) an Markov with a transient probability density

$$\begin{aligned} f_i(y_n/y_{n-1}^{n-p}, n) &= \frac{1}{\sqrt{2\pi}B(y_{n-1}^{n-p}, n)} \\ &\times \exp\left\{-(y_n - A_m(y_{n-1}^{n-p}, n))^2/(2B^2)\right\}. \end{aligned} \tag{20}$$

Consequently, we shall get the current or pointwise classification of process (12), substituting density (20) to recurrent equations (18) or (19).

In the case of Markov processes defined by autoregression equations (13), the likelihood function satisfies the recurrent equation

$$L_m(n) = \frac{1}{\sqrt{2\pi}b_m} \tag{21}$$

$$\times \exp\left\{-(y_n - a_0^{(m)} - \sum_{r=1}^{p} a_r^{(m)} y_{n-r})^2/(2b_m^2)\right\} l_m(n-1).$$

Talking into consideration the normality of transient probability density in (13) for the likelihood function logarithm, we get the recurrent equation

$$l_m(n) = l_m(n-1) - (y_n - a_0^{(m)} - \sum_{r=1}^{p} a_r^{(m)} y_{n-r})^2 \tag{22}$$

$$/(2b_m^2) - \ln b_m - \frac{1}{2}\ln 2\pi$$

with the initial conditions

$$l_m(p) = \ln f_m(y_p^1).$$

Decision rules (9), (10) and the recurrent equations for sufficient statistics (18), (19), (21), (22) allow us to obtain dynamic classification algorithms of stationary processes in a current time.

7. Group classification in the recognition of non-stationary sequences. Consider some constructive approaches to the description of non-stationary processes on the basis of stationary models (13).

Let us call the stationary process y_n, $0 \leqslant n \leqslant N$ a piecewise-stationary, if its transient probability distribution density is piecewise-constant and of the shape

$$f(y_n/y_{n-1}^{n-p}, n) = \begin{cases} f_{m_1}(y_n/y_{n-1}^{n-p}) & \text{if } 0 \leqslant n \leqslant t_1, \\ f_{m_2}(y_n/y_{n-1}^{n-p}) & \text{if } t_1 \leqslant n \leqslant t_2, \\ \cdots & \cdots \\ f_{m_\mu}(y_n/y_{n-1}^{n-p}) & \text{if } t_{\mu-1} \leqslant n \leqslant t_\mu = N, \end{cases} \tag{23}$$

where m_i takes the value from the set $\overline{1, K_m}$, K_m is the number of subclasses $h_{m,k}$ of class h_m. For transient probabilities $f_{m_i}(y_n/y_{n-1}^{n-p})$ the conditions of stationarity of their corresponding processes are accomplished, the number of switching points t_i equals to $\mu - 1$, $\mu \ll N$ and $t_i - t_{i-1} \geqslant T$, $i = \overline{1, \mu}$. The choice T ensures the set length of stationary segments.

Let us also consider the class of locally-stationary processes, possessing such a property that the segments of these processes considered in small intervals of length T can be certainly considered as the realizations of stationary processes.

The approach under consideration aimed at the recognition of piecewise-stationary and locally-stationary process consists of:

– partitioning of non-stationary realizations of class h_m into the segments of stationary observations;

– construction of subclasses $h_{m,k}$ of the stationary sequences y_n for each class h_m;

– description of subclasses $h_{m,k}$ by AR models which are standarts of these classes;

– classification of stationary segments of the observed realization;

– recognition of a non-stationary process realization.

For the partitioning of a non-stationary realization into stationary segments the aposteriori or sequential methods and segmentation algorithms are used, based on a detection of a change in properties of piecewise-stationary processes or a partitioning of locally-stationary sequences into stationary segments. The construction of subclasses of stationary segments $h_{m,k}$ is equivalent to the clustering problem and is realized in the learning regime. It is possible to carry out the clustering both in the space of the stationary segments realization processes and in the space of the parameters, calculated for each segment.

The method, based on the construction of subclasses, al-

lows us to obtain a constructive description of the classes of non-stationary processes and to construct the decision rules for group classifiers. Under this approach the number of classes significantly increases and a necessity for the recognition of realization segments of the determined length arises. The use of recurrent equations (22) for group classification as well as of those received on the basis of their expressions for sufficient statistics requires much time and is not advisable.

From (22) we get the expression for the likelihood function logarithm of the realization y_n of the length T

$$l_m = -T \ln b_m + \frac{1}{2} \ln |M_p^{(m)}| - \frac{1}{2b_m^2} (\bar{y}_p^1)' M_p^{(m)} (\bar{y}_p^1) - \frac{1}{2b_m^2} \sum_{n=p+1}^{T} \left(\bar{y}_n - \sum_{r=1}^{p} a_r^{(m)} \bar{y}_{n-r} \right)^2, \tag{24}$$

where

$$M_p^{(m)} = \Gamma_m^{-1} b_m^2, \ \bar{y}_n = y_n - \mu_n, \quad \Gamma_m = E[y_p^1 \cdot (y_p^1)'].$$

Expression (24) as well as recurrent equation (22) for $l_m(n)$ requires a pointwise calculation of the observed realization by call the models of the class. For group classification algorithms let us transform (24) to the shape

$$l_m(n) = -T \ln b_m + \frac{1}{2} \ln |M_p^{(m)}| - \frac{1}{2b_m^2} (\bar{y}_p^1)' M_p^{(m)} (\bar{y}_p^1) - \frac{1}{2b_m^2} A_m' \sum_{n=p+1}^{T} (\bar{y}_n^{n-p})(\bar{y}_n^{n-p})' A_m, \tag{25}$$

where the vector

$$A_m = (-a_p^{(m)}, -a_{p-1}^{(m)}, \ldots - a_1^{(m)}, 1).$$

Let us present the quadratic form

$$Q_m = (\bar{y}_p^1)' M_p^{(m)} (\bar{y}_p^1)$$

in the shape

$$Q_m = A_m' S_p A_m.$$

The matrix S_p is defined the help of the methods, presented in Box, Jenkins (1970). Denoting the matrix $\sum_{n=p+1}^{T} (y_n^{n-p})(y_n^{n-p})'$ by R_p, we get for (25) the expression

$$l_m = -T \ln b_m + \frac{1}{2} \ln |M_p| - \frac{1}{2b_m^2} A_m' G_p A_m, \qquad (26)$$

where

$$G_p = S_p + R_p.$$

The elements of the symmetric matrix $G_p = \{g_{ij}\}$ of the size $[p+1, p+1]$ are equal to

$$g_{ij} = \sum_{n=p+3-i-j}^{T-p} \bar{y}_{n+i-1}\, \bar{y}_{n+j-1}. \qquad (27)$$

The determinant $|M_p|$ may be expressed by means of parameters of the AR model, using statistics (26) for the classification of segments of the realizations T' readings long.

For the use of statistics in group classification algorithms it is advisable to select the mean value μ_m. Substituting $\bar{y}_n = y_n - \mu_m$, we get

$$\begin{aligned} g_{ij} &= \sum_{n=p+3-i-j}^{T-p} (y_{n+i-1} - \mu_m)(y_{n+j-1} - \mu_m) \\ &= \sum_{n=p+3-i-j}^{T-p} y_{n+i-1}\, y_{n+j-1} \qquad (28) \\ &\quad - \mu_m \sum_{n=p+3-i-j}^{T-p} (y_{p+i-1} + y_{p+j-1}) + \mu_m^2 D_3. \end{aligned}$$

Having introduced the matrixes $D_1 = \{d^1_{ij}\}$, $D_2 = \{d^2_{ij}\}$, $D_3 = \{d^3_{ij}\}$
with the elements

$$\begin{aligned} d^1_{ij} &= \sum_{n=p+3-i-j}^{T-p} y_{n+i-1}\, y_{n+j-1}, \\ d^2_{ij} &= \sum_{n=p+3-i-j}^{T-p} (y_{n+i-1} + y_{n+j-1}), \\ d^3_{ij} &= T - 2(p+1) + i + j, \end{aligned} \tag{29}$$

we get the final expression for the statistics

$$\begin{aligned} l_m &= -T \ln b_m + \frac{1}{2} \ln |M_n| \\ &- \frac{l}{2b^2_m} [A'_m D_1 A_m - \mu_m A'_m D_2 A_m + \mu^2_m A^1_m D_3 A_m]. \end{aligned} \tag{30}$$

It is convenient to use statistics l_m (30) for the realization of group classification at a large number of classes. After the calculation it is easy to determine the statistics: the likelihood function, the likelihood ratio logarithm

$$L_m = \exp\{l_m\}, \; u_{ij} = l_i - l_j$$

the aposteriori probability of the class

$$p_m(N) = \frac{p_m L_m}{\sum_{m=1}^{M} p_m L_m}.$$

The solution is made on the basis of decision rule (10). The use of expressions (30) allows us to significantly diminish computer time.

8. Comparison of algorithms of pointwise and group classification. The dynamic classification algorithms which use recurrent equations for likelihood function logarithm (22) or corresponding expressions (24), require a piecewise calculation of the observed realization by each cluster, i.e., expressions (24) are calculated consequently by realization readings for each AR model. Group classification algorithms also require the calculation of statistics (30) for each cluster, however entering in (30) the matrices D_1, D_2, D_3 do not depend on the cluster (class) and are calculated by observed realization (6) once in the process of its recognition. Since algorithms (24) and (30) use the identical statistics $l_{m,k}$, their comparison may be carried out by the expenditures of computer time. In this connection it should be noted that the matrices D_1, D_2, D_3 in expression (30) are symmetric

$$d_{ij}^l = d_{ji}^l, \quad l = 1,2,3, \quad i,j = \overline{1,p+1},$$

and the elements of the main diagonal and parallel to it may be calculated recurrently

$$\begin{aligned} d_{i+1,j+1}^1 &= d_{ij}^1 + y_{p+1-j}y_{p+1-j} + y_{T-p+i}y_{T-p+j}, \\ d_{i+1,j+1}^2 &= d_{ij}^2 + y_{p+1-j}y_{p+1-j} + y_{T-p+i}y_{T-p+j}. \end{aligned} \tag{31}$$

For the given AR order equal to p, the number of classes M and the length of the observed realization equal to T, the number of multiplication and division operations in current classification (24) is equal to

$$yD_1 \approx (p+2)MT, \tag{32}$$

and the group of classification (30)

$$yD_2 \approx (p+1)T + (p+1)p + 3(p+1)(p+2)M + 3M. \tag{33}$$

A similar number of compositions, substractions and sendings is equal to

$$CB_1 \approx [(p+3)M + 2M]T,$$
$$CB_2 \approx [2(p+1)T + 2(p+1)p + 3(p+1)(p+2)M + 4M],$$

the number of indexes

$$UH_1 \approx pMT,$$
$$UH_2 \approx (2p+3)(T+p) + p.$$

From the given calculations it is obvious that for a large number of readings ($T = 1000$) group classification algorithms gain in time approximately M times. Equating to (32) and (33), it is possible to get the smallest number of readings $T_{\min}$, at which the number of multiplication and division operations both in current and group classifications in the same:

$$T_{\min} = \frac{3(p+1)(p+2)M + (p+1)p + 3M}{(p+2)M - (p+1)}.$$

9. Clustering in group classification of random sequences. The construction of subclasses of the stationary segments in group classification of non-stationary processes is carried out in the learning regime with the use of clustering algorithms. The main parameters of clustering algorithms are the following: the distance measure between the objects of a set and threshold constants defining the solution in the very clustering process. For the construction of clusters in the space of realizations the aposteriori probability of a cluster, the likelihood function, the likelihood ratio, the Itakura–Saito distance, ect. are used as distance measures. The Euclidean distance is usually used for clustering in the space of parameters. As a result of clustering an initial set of objects must

be divided into clusters and a standart for each cluster must be chosen. For complex processes in the case of a great order AR (Montvilas, 1987) it is necessary to build complex hypersurfaces for the construction of clusters. A constructive clustering method is the method when a nonlinear mapping of the signal onto the plane in the shape of a two-dimensional vector is used. Its essence is in the following. Each stationary segment y_{n-T}^{n} of the considered non-stationary process may be described by autoregression model (13) and presented by $L = p + 2$-dimensional vector with the characterizing each vector parameters of AR: $m, b, a_k (k = 1, \ldots p)$. Let us denote these vectors x_i, $i = 1, 2, \ldots, R$; where R is the number of stationary segments. We shall map the vectors x_i onto the plane by two-dimensional vectors z_i, $i = 1, 2, \ldots, R$. The main requirement for mapping of L-dimensional vectors by 2-dimensional vectors is to retain the inner structure of distances in the hyperspace between L-dimensional vectors after mapping them on two-dimensional vectors. This is achieved using the method of nonlinear mapping (Sammom, 1969).

In practical cases it is possible to realize the clustering not only after the receipt of all the information on a random process (Montvilas, Tshyjov, 1989), but also in a current time after an additional inflow of information on S stationary segments of the random process under clustering, in this connection the expressions of nonlinear mapping transform respectively.

Thus, let us have in L-dimensional space $R + S$ vectors: x_i, $i = 1, \ldots, R$; x_j, $j = R + 1, \ldots, R + S$. R vectors are already mapped on 2-dimensional vectors z_i, $i = 1, \ldots, R$. We shall map on the plane the S vectors z_j, $j = R+1, \ldots, R+S$. In the capacity of the initial conditions the vectors z_j, $j = R + 1, \ldots, R + S$, in a random way arrange themselves on the plane within the limits of R vector reflected values. Let us denote the distances between the vectors x_i and x_j in L-

dimensional space by d^x_{ij}, while by d^z_{ij} on the plane, respectively. In the case of absence some kind of the apriori information on the investigated random process Euclidean metric is used as a measure. After that the normalized mean value of errors in distances denoted by H is calculated. Besides for the solution of a nonlinear mapping problem it is necessary to change the location of the vectors z_j, $j = R+1, \ldots, R+S$ on the plane in such a way that the error H would be minimal. It is realized by the method of a steepest descent. After the n-th iteration the distance error will be

$$H(n) = \frac{1}{\sum\limits_{i=1}^{R}\sum\limits_{j=R+1}^{R+S} d^x_{ij}} \sum_{i=1}^{R}\sum_{j=R+1}^{R+S} [d^x_{ij} - d^z_{ij}(n)]^2 / d^x_{ij} \qquad (34)$$

in this connection

$$d^z_{ij}(n) = \sqrt{\sum_{k=1}^{2} [z_{ik} - z_{jk}(n)]^2}, \qquad (35)$$
$$i = \overline{1, R};\ j = \overline{R+1, R+S}.$$

For $n+1$-th iteration the coordinates of the mapped vectors z_j will be

$$z_{jq}(n+1) = z_{jq}(n) - F\Delta_{jq}(n), \qquad (36)$$
$$j = R+1, \ldots, R+S,\ q = 1, 2;$$

where

$$\Delta_{jq}(n) = \frac{\partial H(n)}{\partial z_{jq}(n)} \Big/ \left| \frac{\partial^2 H(n)}{\partial z^2_{jq}(n)} \right|, \qquad (37)$$

F-coefficient for the correction of coordinates is defined empirically, $F = 0.35$;

$$\frac{\partial H}{\partial z_{jq}} = E \sum_{i=1}^{R} \frac{D \cdot B}{d^x_{ij}\, d^z_{ij}}, \qquad (38)$$

$$\frac{\partial^2 H}{\partial z_{jq}^2} = E \sum_{i=1}^{R} \frac{1}{d_{ij}^x d_{ij}^z} [D - \frac{B^2}{d_{ij}^z}(1 + \frac{D}{d_{ij}^z})], \qquad (39)$$

where

$$E = -\frac{2}{\sum_{i=1}^{R} d_{ij}^x}, \quad D = d_{ij}^x - d_{ij}^z, \quad B = z_{jq} - z_{iq}.$$

It is possible to realize a sequential clustering of piecewise-stationary multiparametric random processes after the inflow of information on each current stationary segment of the random process. Then for nonlinear mapping the above presented expressions for $S = 1$ are used.

It should be noted that with regard to the latter remark the clustering method, based on nonlinear mapping of the classified stationary segments onto the plane, may be successfully applied for a sequential detection of many changes in several unknown states of the dynamic systems, generating multiparametric random processes. In this connection in order to follow the changes, e.g. the states of the technological process and its recognition, it is very convenient to reproduce the state of the process by a mark on the screen of PC and, having in mind the existence of particular states, to identify the current state, a deviation from it or a transfer to another state.

The developed methods and clustering, piecewise and group classification algorithms are practically realized by a computer in the shape of a program package and applied for the solution of practical problems.

Conclusions

1. The methods of dynamic recognition, based on the use of stochastic dynamic models for the description of classes allow us to obtain the algorithms of piecewise (current) classification and those of group classification of the sequences.

2. Piecewise classification algorithms allow us to develop the methods of sequential multialternative recognition and to construct a multichanuel recognizing device of conveyer type.

3. For the recognition of non-stationary piecewise or locally-stationary sequences it is advisable to use group classification algorithms, based on the construction of the subclasses of stationary segments of the recognizable sequence. The use of group classification algorithms allows us to reduce the recognition time for a large number of subclasses and a big length of recognizable segments.

4. For the construction of subclasses and the choice of standarts it is advisable to use clustering methods, based on nonlinear mapping of the classified segments onto the plane, besides it is possible to use these methods for a sequential detection of many changes in several unknown properties of random processes.

REFERENCES

Atamukas, M.S., and E.K.Shpilewski (1979). Adaptive approach in the recognition of nonstationary processes. In E.Shpilewski (Ed.), *Statisticheskie Problemy Upravleniya,* Vol.37. Inst. Math. Cybern. Lith. Acad. Sci., Vilnius. pp. 83–93 (in Russian).

Box, G., and G.Jenkis (1970). *Time series analysis, Forecasting and Control.* San Francisco.

Krutowski, A.V., and E.K.Shpilewski. Multialternative recognition, detection and estimation of change points in the properties of composite stochastic processes by sequential observations. In E.Shpilewski (Ed.), *Statisticheskie Problemy Upravleniya,* Vol.69. Inst. Math. Cybern. Lith. Acad. Sci., Vilnius. pp. 47–57 (in Russian).

Liptser, R.Sh., and A.N.Shiryaev (1974). *Statistic of Stochastic Processes.* Nauka, Moscow (in Russian).

Montvilas, A.-M.M. (1987). Speech compression by variable length of the frame. In E.Shpilewski (Ed.), *Statisticheskie Problemy Upravleniya,* Vol.78. Inst. Math. Cybern. Lith. Acad. Sci., Vilnius. pp. 23–31 (in Russian).

Montvilas, A.-M.M., and A.P.Tshyjov (1989). A mapping of the manyparametric nonstationary clustering processes. In E.Shpilewski (Ed.), *Statisticheskie Problemy Upravleniya,* Vol.85. Inst. Math. Cybern. Lith. Acad. Sci., Vilnius. pp. 80–92 (in Russian).

Petrov, G.J., and E.K.Shpilevski (1974). Some problems in statistic of random processes in dynamic systems. In J.Tsypkin (Ed.), *VI Vsesojuznoe Soveschanie po Problemam Upravleniya,* Part I. Moscow. pp. 181–184 (in Russian).

Sammon John.W. (1969). A Nonlinear mapping for data structure analysis. *IEEE Trans. on Computers,* **18**(5), 401–409 .

Shpilewski, E.K. (1971). Optimal classification of observations of random processes. In L.Telksnys (Ed.), *Statisticheskie Problemy Upravleniya,* Vol.1. Inst. Math. Cybern. Lith. Acad. Sci., Vilnius. pp. 61–75 (in Russian).

Shpilewski, E.K. (1980). Dynamic classification of stochastic processes and systems in discret time. *Avtomatika i telemechanika,* **11**, 46–52; **12**, 45–54 (in Russian).

Shpilewski, E.K. (1986). Sequential multialternative recognition, detection and estimation of change-points in the structure of dynamic systems. In N.K.Sinha and L.A.Telksnys (Eds.), *2nd IFAC Symposium on stochastic control,* part II, Vilnius. IFAC, Moscow. pp. 219–223 .

Wald, A. (1946). Statistical lecision function. *Ann. Math. Stat.,* **20.**

Received January 1990

A.–M.Montvilas is a Candidate of Technical Sciences, a senior researcher at the Process Recognition Department of the Institute of Mathematics and Informatics, Lith. Acad. Sci. Scientific interests include: processing of random processes, classification, detection of change in the properties of random processes, simulation.

R.Narkowich is an engineer of Vilnius Scientific Research Institute. Scientific interests include: mathematical simulation, application of computers, microelectronics.

E.Shpilewski is a Doctor of Technical Sciences, a principal scientific collaborator of the Institute of Mathematics and Informatics, Lith. Acad. Sci. Scientific interests include: mathematical statistics and simulation, application of computers, cybernetics, pattern recognition.