Classification of Seismic Signals of a Poorly Studied Nature

A. V. Timofeev

Research Institute for Automation and Electromechanics, ul. Belinskogo 53, Tomsk, Russia

Abstract – This article dwells upon a method of classification of broadband stochastic signals that can be easily implemented and is fairly efficient. Seismic signals of relaxation phenomena in the region of open-pit mining are considered as signals of this type.

1. INTRODUCTION

This article is devoted to the description of an efficient algorithm of classification of stochastic signals of a poorly studied nature. This algorithm implements a procedure of learning with a teacher, which gives classification of stochastic signals on the basis of implicitly formulated criteria. An expert that is able to classify signals with a high percentage of correct outcomes on the basis of poorly formalized information that implicitly characterizes the physical essence of detected signals, plays the role of the teacher. In this case, the method proposed allowed one to effectively imitate actions of this expert. The recognition system constructed on the basis of the principle presented in the article has the ability to operate in real time. The presentation is based on an example of solving the problem of recognizing seismic signals entering a system of spatially distributed seismic pickup units that are parts of a system that ensures security of underground works in a region of open-pit mining. The system is suited for tracing technological discipline in exploiting open-pit mining, as well as for tracing and timely discovery of catastrophic change of state of shock hazards in the region of mining. To solve these problems, it is necessary to solve an additional problem of recognition of types of seismic signals and noise entering the registration system. The system of classification of stochastic signals described in this paper functions as a part of the system of security support for underground works in the region of open-pit mining of the Komsomolskii mine (the Noril'sk seismic station, Noril'sk industrial region).

2. SYSTEM OF NOTATION AND DEFINITIONS

We introduce the following notation:

S is the set of objects to be classified;

S_E is the set of objects presented for examination, $S_E \subseteq S$

$S^{(i)}$ is the set of objects that belong to the ith class,

$$S^{(i)} \subseteq S; \bigcup_{i=1}^{t} S^{(i)} = S, \text{ and } \bigcap_{i=1}^{t} S^{(i)} = \{\varnothing\}.$$

t is the number of classes of phenomena set by an expert *a priori*.

In practice, the object of observation is usually not observable. Therefore, the study of this object is carried out on the basis of its actual measurable images. In our case, a set of multidimensional frequency–temporal seismic signals accepted by a system of spaced seismic sensors were analyzed as such images. For a signal, a scalar function of its image is called its partial formal attribute. Assume that there exist m such functions. In this case, a particular object corresponds to an m-dimensional vector whose elements are numerical values of partial formal attributes of this object. Every ith element (the ith formal attribute) of this vector is measurable in the corresponding metric space $\mathbf{D}(i)$. We denote the entire space of formal attributes by P^m, where $P^m = \prod_{i=1}^{m} (d) \mathbf{D}(i)$ and $\prod(d)$ denotes the Cartesian product.

We introduce the following notation:

$i(s)$ is the image of an object $s \in \mathbf{S}$;

$\mathbf{I}$ is the set of possible images of objects $s \in S$;

$\tilde{i}(s)$ is a graphic representation of $i(s) \in \mathbf{I}$;

$\mathbf{f}(i(s))$ is the vector of formal attributes of the object $s \in S$ computed by the image $i(s)$,

$$\forall S \in S \ \exists \ \mathbf{f}(i(s)) \in P^m$$

F is the set of admissible vectors of formal attributes of objects,

$$\mathbf{F} = \{\mathbf{f} \in P^m | \ \forall \mathbf{f} \in F \ \exists \ (i(s) \in \mathbf{I}, s \in S): \mathbf{f} = \mathbf{f}(i(s))\}; (1)$$

$\mu(x, A)$ is the membership function for an element x and the set A;

$\rho_i(x, y)$ is a metric in $\mathbf{D}(i)$; $x, y \in \mathbf{D}(i)$.

Elements of the attribute space that correspond to certain objects from S are randomly implemented in the process of observation and are measured in the Borel probability space $Q = (P^m, \Omega, \mathbf{P})$. Moreover, let

$\mathbf{I_E}$ denote the set of images of objects to be examined;

$\mathbf{F_E}$ denote the set of vectors of admissible formal attributes of objects from S_E depicted by elements of $\mathbf{I_E}$;

$\mathbf{M}(t)$ be the matrix of expert estimates; the columns of this matrix are vectors $\langle \mu(\tilde{s}_i, S^{(1)}) \ \mu(\tilde{s}_i, S^{(2)}) \ ... \ \mu(\tilde{s}_i, S^{(t)}) \rangle^{\mathsf{T}}$, where the quantity $\mu(\tilde{s}_i, S^{(k)})$ denotes the membership

$$\{S_E\} \xrightarrow{\;q\;} \{I_E\} \xrightarrow{\;f\;} \{F_E\} \xrightarrow{\;\Xi\;} \{M(t)\}$$
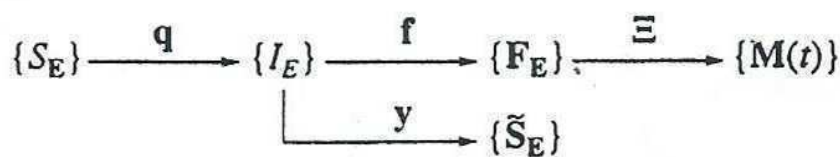$$\qquad\qquad\;\; \xrightarrow{\;y\;} \{\tilde{S}_E\}$$

Fig. 1.

degree for an object $s_i \in S_E$ having the graphic image $\tilde{S}_i$ with respect to the set $S^{(k)}$, or in other words, with respect to the kth class of phenomena;

$\tilde{S}_E$ be the set of graphic images of elements of the set I_E.

3. STATEMENT OF THE PROBLEM

Consider the situation when, for any formal attribute f that is a component of the attribute vector $\mathbf{f}$, an *a priori* expert estimate of the amount of resources (computational, material, etc.) required for using this attribute in solving classification problems is known. We denote this value by the symbol $C(f)$, and by $T(\mathbf{f})$, we denote the sum of the values of $C(\cdot)$ that correspond to a particular vector f. Obviously, $T(\mathbf{f})$ is equal to the amount of resources needed to implement the recognizing procedure that uses the vector $\mathbf{f}$. The value G equal to the maximally admissible amount of resources needed to implement the procedure of automatic classification is known.

Observing images $i(s) \in \mathbf{I}$ of classification objects $s \in \mathbf{S}$ with a matrix $\mathbf{M}(t)$ of expert estimates at hand, we need to solve the following problems:

(1) construct a procedure of automatic classification of objects $s \in S$ and propose estimates of the quality of classification solutions,

(2) construct an effective procedure for estimating the vector

$$\langle \mu(s, S^{(1)})\ \mu(s, S^{(2)}) \dots \dots \mu(s, S^{(t)}) \rangle^T$$

for an arbitrary $s \in \mathbf{S}$,

(3) propose an optimality criterion for solution of problems (1) and (2),

(4) choose a vector of formal attributes $\mathbf{f}^*(\cdot)$ such that it allows one to obtain a solution to problems (1) and (2) that is optimal in the sense of the chosen optimality criterion and obeys the inequality $T(\mathbf{f}^*) \leq G$,

(5) propose particular metric spaces where formal attributes for automatic classification of seismic signals can be measured.

Problems (1) - (5) should be solved under the condition of the absence of *a priori* information on the degree of mutual dependence between the values of particular formal attributes and the number of the class to which the object under study actually belongs. In doing this, we should take into account the fact that the attributes of objects are possibly dependent both in the probabilistic and in the usual sense. However, *a priori* information on these dependences is also absent. In the

case considered, the objects of classification are seismic signals whose sources are natural or artificial phenomena (rock shocks, processes of spontaneous cracking of rocks, technological noise, etc.)

4. LEARNING OF THE SYSTEM OF AUTOMATIC CLASSIFICATION

Consider a situation when stochastic signals coming from sources of a poorly explored nature are observed within the framework of a monitoring system. The images of these signals constructed in various coordinate systems are visually observable. An expert–operator has great experience in classifying observed signals using their images. This fact permits *a priori* determination of the number of classes of sources of these signals. Naturally, since expert–operators are usually not specialists in the domain of pattern recognition, they are not able to give a formal means of automatic classification of observable signals. The operator is only able to construct a training sample that carries information concerning the correspondence of visually processed signals to some of their formal attributes, the set of which is formed in advance by designers of the classification system. The choice of the optimal set of formal attributes for describing the objects of automatic classification is a separate problem, whose solution will be proposed in our next paper. We note in advance that the designers of the classification system are not specialists in the visual classification of seismic signals. However, they have an *a priori* opportunity to form an admissible vector of formal attributes of classified signals. This vector is usually redundant for the solution to the problem stated. However, special methods for diminishing the redundancy of this vector do exist. One of these methods will be described below. Components of this *a priori* formed vector are certain real functions defined for its digitized realization (spectral features, amplitude features, etc.).

Consider a concise general scheme of processing information at the stage of pattern-recognition learning. A diagram of this procedure is presented in Fig. 1. Here, the function $\mathbf{q}$ maps the set of objects S_E to be examined by the expert onto the set of images of these objects I_E. The function $\mathbf{f}$ maps the set I_E onto the set of admissible formal attributes F_E. The function $\mathbf{y}$ maps the set I_E onto the set $\tilde{S}_E$ that consists of graphic images of the elements of the set I_E. The function Ξ maps the set of formal attributes F_E onto the set of vectors of expert estimates that constitute the matrix $\mathbf{M}(t)$. This matrix has the following form:

$$\mathbf{M}(t) = \begin{vmatrix} \mu(\tilde{s}_1, S^{(1)}) & \mu(\tilde{s}_2, S^{(1)}) & \dots & \mu(\tilde{s}_z, S^{(1)}) \\ \mu(\tilde{s}_1, S^{(2)}) & \mu(\tilde{s}_2, S^{(2)}) & \dots & \mu(\tilde{s}_z, S^{(2)}) \\ \vdots & & & \\ \mu(\tilde{s}_1, S^{(t)}) & \dots & \dots & \mu(\tilde{s}_z, S^{(t)}) \end{vmatrix}.$$

Here, the quantity z denotes the cardinal number of the set S_E. The expert forms the matrix $M(t)$ by analyzing graphic representation of images of several objects found *a priori* by the system of registration of seismic signals. If the physical nature of these objects is known, for example, if it is *a priori* known that the graphic image of a technological explosion is analyzed, the expert should form a column of the form

$$\langle M(t) \rangle_i = \langle 0, 0, 0, 0, ..., 1, 0, 0, 0 \rangle^T$$

in this matrix. Here, unity corresponds to the class of objects of the type "technological explosion" and i is the number of the column of the matrix $M(t)$ equal to the number of the group of images of the object analyzed. In the case when the source of registered signals is not known in advance, the expert forms a column of the form

$$\langle M(t) \rangle_i = \langle \ \mu(\tilde{s}_i, S^{(1)}) \ \mu(\tilde{s}_i, S^{(2)}) \ ... \ \mu(\tilde{s}_i, S^{(t)}) \ \rangle^T,$$

where the conditions

$$\mu(\tilde{s}_i, S^{(j)}) \in \ ]0, 1[\ , \qquad (2)$$

$$\sum_j \mu(\tilde{s}_i, S^{(j)}) = 1 \qquad (3)$$

hold. Thus, in this case, the expert indicates the degree to which the group of object images presented for examination corresponds to characteristics of each of t *a priori* possible classes of phenomena. The correspondence degree of the images of the ith object to the characteristics of the jth class of phenomena is estimated by the value $\mu(\tilde{s}_i, S^{(j)})$, i.e., by the membership degree of the element s_i with respect to the set $S^{(j)}$. To make the input of expert estimates easier, a special program shell was designed, which allows one to fix membership coefficients of an object with respect to any of t classes in a convenient graphic form while automatically tracing that conditions (2) and (3) are satisfied. The expert can observe the graphic image of this object on the display.

Consider an arbitrary element $\mu(\tilde{s}, S^{(k)})$ of the matrix $M(t)$, which is by definition the degree of membership of the object s in the class S^k and reflects the subjective opinion of the expert on the conformity of the graphic images of the object to be examined to the features of the kth class of phenomena. As was mentioned above, every object s can be described not only by its complete digital images but also by a vector of formal attributes $f(s)$. Obviously, since a description of this type is more compact, it implies loss of useful information on object images. However, it is this description that is convenient for processing and storing the information about objects being analyzed in the computer memory. Eventually, by definition, the images of objects do not contain complete information on the essence of phenomena that generated them.
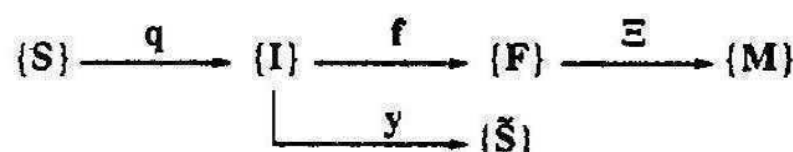

Fig. 2.

Thus, the following set is admissible for automatic processing after examination:

$$H = \{ \ \langle f(s_i), \langle M(t) \rangle_i \rangle \ | \ (s_i \in S_E) \ \},$$

where

$$\langle M(t) \rangle_i = \langle \mu(\tilde{s}_i, S^{(1)}) \ \mu(\tilde{s}_i, S^{(2)}) \ ... \ \mu(\tilde{s}_i, S^{(t)}) \rangle^T.$$

It is convenient to consider the function Ξ as a function that maps the set F onto the set

$$M = \{ \ \langle \mu(\tilde{s}, S^{(1)}), \mu(\tilde{s}, S^{(2)}), ...\mu(\tilde{s}, S^{(t)}) \rangle^T \ | \ s \in S \}$$

and that is given only by a set of support points H. Formally, we can consider the stage of learning complete after formation of this set.

5. ALGORITHM OF AUTOMATIC CLASSIFICATION

At the stage of automatic classification, it is necessary to solve the problem of reconstruction of the function Ξ from the set of data that forms the set H. To solve this problem, we used the method of kernel local approximation [1]. However, before we proceed to its description, consider briefly the scheme of information processing at the stage of automatic classification. A diagram of this procedure is presented in Fig. 2. Here I is the set of all possible images of objects from S,

F is the set of admissible vectors of formal attributes of objects from S,

$\tilde{S}$ is the set of graphic images of elements of I,

M is the set of possible expert estimates that correspond to the sets $\tilde{S}$ and F.

The functions q, f, and y have, generally speaking, the same meanings as before. The function $\tilde{\Xi}$ is an approximation of the function Ξ. In this case, a seismic signal $i(s) \in I$, which is an image of a phenomenon $s \in I$, enters the system of preprocessing for obtaining the vector of formal attributes $f(i(s)) \in F$ that correspond to this signal. This signal is simultaneously displayed on the control graphic monitor in the form $y(i(s) \in \tilde{S}$. Next, starting from the value of the vector $f(i(s))$ and using the function $\tilde{\Xi}$, we compute estimates of the membership degrees of the identified phenomenon with respect to each of t *a priori* defined classes of phenomena. A solution is obtained in the form of a t-dimensional vector with real components. A phenomenon is considered to belong to the ith class if the ith

component of this vector is the greatest with respect to the remaining $(t-1)$ components.

As was mentioned above, we used the method of nonparametric local approximating the kernel type [1] for approximation of the function $\Xi(\cdot)$. We schematically describe the methodology of application of this technique for the case under study.

Consider the vectors

$$\mathbf{T}(x) = (T(x, \mathbf{f}_1), T(x, \mathbf{f}_2), ..., T(x, \mathbf{f}_t)) \in R^t$$

$$\mathbf{f}_i, x \in R^m, \quad i \in \{1, ..., m\},$$

$$T(\mathbf{x}, \mathbf{y}) = \prod_{k=1}^{m} \mathbf{J} [\rho_k(\langle \mathbf{x} \rangle_k, \langle \mathbf{y} \rangle_k) \delta_k^{-1}]; \quad \mathbf{y}, \mathbf{x} \in R^m. \quad (4)$$

Here, $\Delta = (\delta_1, \delta_2, ... \delta_m) \in R^m$ is a parameter of locality; the quantities $\mathbf{x}, \mathbf{f}_j \in \mathbf{F}$, where $\mathbf{f}_j \in \mathbf{F_E}$, are measurable in the probability space; the record $\langle \mathbf{a} \rangle_i$ denotes the ith component of the vector $\mathbf{a}$; $\mathbf{x}$ is the vector of formal parameters of the processed signals, and $\mathbf{J}(\cdot)$ is a bell-shaped, square-integrable function whose particular form will be defined below.

Following the method of nonparametric local kernel approximation, we use the function $\tilde{\Xi}_z(\mathbf{f}, \Delta) \in R^m$, $\mathbf{f} \in \mathbf{F}$ as an approximation of the function $\Xi(s)$. The former function is defined as follows:

$$\tilde{\Xi}_z(\mathbf{f}, \Delta) = \mathbf{M}(t)\mathbf{T}^T(\mathbf{f}) [\mathbf{T}(x)\epsilon_z^T]^{-1},$$

$$\epsilon_z^T = (1, 1, ..., 1) \in R^t.$$

The function $\mathbf{J}(\cdot)$ is used as a kernel function $\mathbf{J}(x) = \exp(-|x|)$.

Choice of approximation parameters. Consider the functional $Y_z(\Delta)$ that characterizes the precision of the model $\tilde{\Xi}_z(\mathbf{f}, \Delta)$ according to its errors in the set of support points $\mathbf{H}$:

$$Y_z(\Delta) = \sum_{f_i \in \mathbf{F_E}} \left\| \tilde{\Xi}_z(\mathbf{f}, \Delta) - \mathbf{M}(t) \right\|_v.$$

Here, $\|\cdot\|_v$ denotes the matrix norm of the matrix. The function $Y_z(\Delta)$ features interpolation properties of the estimate $\tilde{\Xi}_z(\mathbf{f}, \Delta)$ and makes it possible to apply this estimate for the values $\mathbf{f} \notin \mathbf{F_E}$, i.e., values that are distinct from the nodes of the set $\mathbf{H}$ (which gives rise to the estimate). In this case, the optimal value of the locality parameter Δ^* is obviously defined in the following way:

$$\Delta^* = \arg \inf_{\Delta \in R^m} Y_z(\Delta).$$

Since particular formal attributes may belong to various metric spaces or have various physical dimensions, to use the method of nonparametric local approx-

imation, we should perform a special scaling of formal attributes. We assume that

$$\overset{m}{\underset{i=1}{\forall}} \exists q_1^i, \quad q_2^i \in R^1 : \forall \mathbf{f} \in \mathbf{F}: \langle \mathbf{f} \rangle_i \in [q_1^i, q_2^i].$$

In this case, following [1], the procedure of scaling will consist in the substitution of the quantities $R_i(\langle \mathbf{x} \rangle_i)$ and $R_i(\langle \mathbf{y} \rangle_i)$ for the quantities $\langle \mathbf{x} \rangle_i$ and $\langle \mathbf{y} \rangle_i, i \in \{1, ..., m\}$, in expression (4), respectively. Here,

$$R(s) = [s - 0.5(q_1^i + q_2^i)] [0.5(q_1^i - q_2^i)]^{-1},$$
$$i \in \{1, ..., m\}.$$

Criterion for making decisions concerning signal classification. Let $\mathbf{B}$ denote the set that consists of pairwise-distinct vectors, where the components belong to the set $\mathbf{N} = \{1, 2, 3, ..., N\}$, and the dimension of elements of the set $\mathbf{B}$ does not exceed the value N. Thus,

$$\mathbf{B} = \bigcup_{i=1}^{N} \mathbf{B}'i), \quad \text{where } \mathbf{B}'(i) = \{b \in \mathbf{N}^i | \forall b',$$

$$b'' \in \mathbf{B}(i): \sum_{i,j} \mathfrak{W}(\langle b' \rangle_i, \langle b'' \rangle_j) < i\}$$

$\mathbf{N}^i$ is the i fold Cartesian product of the set $\mathbf{N}$,

$$\mathbf{N} = \{1, 2, 3, ..., N\}, \quad \mathbf{B}(i) = \mathbf{N}^i,$$

$$\mathfrak{W}(b', b'') = \begin{cases} 1, & \text{for } b' = b'' \\ 0, & \text{for } b' \neq b''. \end{cases}$$

Assume that every component of an arbitrary vector $\mathbf{f} \in \mathbf{B}$ denotes a particular formal attribute. Then, by $Y_z(\Delta^*|\mathbf{f})$ we denote the value of the functional $Y_z(\Delta^*)$ constructed for the approximation $\tilde{\Xi}_z(\mathbf{f}, \Delta)$ when using the set of formal attributes that corresponds to the vector $\mathbf{f} \in \mathbf{B}$. The quantity $Y_z(\Delta^*|\mathbf{f})$ characterizes the precision of the approximation of the operator $\Xi(\ )$ obtained on the support set $\mathbf{M}(t)$ when using the vector of attributes $\mathbf{f} \in \mathbf{B}$. Hereafter, if a set of formal attributes $\mathbf{f} \in \mathbf{B}$ is used in the approximation $\tilde{\Xi}_z(\mathbf{f}, \Delta)$, we will write $\tilde{\Xi}_z(\mathbf{f}, \Delta|\mathbf{f})$. We assume that, in the definition of the vector of estimates

$$w(\tilde{s}_i) = (\mu(\tilde{s}_i, S^{(1)}), \mu(\tilde{s}_i, S^{(2)}), ..., \mu(\tilde{s}_i, S^{(t)})),$$

the expert should ensure that a unique maximal component of the vector $w\tilde{s}_i$ exists. This naturally ensures the uniqueness of the classification of the pattern $\tilde{s}_i$ carried out by the expert. We denote this component by $\text{Max}(w\tilde{s}_i)$ and the index of this component by $\text{IND}(\text{Max}(w\tilde{s}_i))$. For every column $\langle \mathbf{M}(t) \rangle_j$ (j is the number of the column) of the matrix $\mathbf{M}(t, U)$ of expert estimates and every column $\langle \tilde{\Xi}_z(\mathbf{f}, \Delta \mid \mathbf{f}) \rangle_j$ of the matrix

$\tilde{\Xi}_z(f, \Delta|f)$, we define the values $\text{IND}(\text{Max}(\langle M(t, U)\rangle_j)$ and $\text{IND}(\text{Max}(\langle \tilde{\Xi}_z(f, \Delta|f)\rangle_j)$, respectively.

We propose the following rule as a criterion for making decisions about the membership of a classified signal with respect to a particular class: the quantity $\text{IND}(\text{Max}(\langle \tilde{\Xi}_z(f(s), \Delta|f)\rangle_j))$ is considered the number of the class to which the classified signal s belongs. Thus, making a decision in accordance with this rule comprises classification.

Estimation of the classification quality. It seems quite natural to use of the values of components of the function $\langle \tilde{\Xi}_z(f(s), \Delta|f)\rangle_j$ for estimating the quality of classification of the signal s. In this case, in accordance with the decision criterion for classification, the value of $(\text{Max}(\langle \tilde{\Xi}_z(f(s), \Delta|f)\rangle_j)$ should be considered an estimate of the membership degree of the signal s with respect to the class $\text{IND}(\text{Max}(\langle \tilde{\Xi}_z(f(s), \cdot)\rangle_j))$. The greater its absolute value, the more reliable the results of the classification should be considered.

Choice of the attribute set for describing the objects to be recognized.

Consider the following real function:

$$\lambda(b) = 100\left(1 - \sum_j |\text{IND}(\text{Max}(\langle M(t, U)\rangle_j))\right.$$

$$\left. - \text{IND}(\text{Max}(\langle \tilde{\Xi}_z(f, \Delta|b)\rangle_j))|/z\right).$$

Here, z is the cardinal number of the set S_E, which consists of signals to be examined. We now determine the meaning of this function. On the set of signals S_E, we define the quantity r equal to the percentage of coincidence of classifications made by experts with classifications obtained using the approximation $\tilde{\Xi}_z(\cdot|f)$. The definition of the function $\lambda(\cdot)$ implies immediately that $r = \lambda(f)$. We can now define the procedure for choosing the vector of the formal attributes f^* that provides the most effective functioning of the classification procedure under *a priori* preset restrictions. Recall that the quantity $Y_z(\Delta^*|f)$ may play the role of a quality index of the reconstruction of the function $\Xi(\cdot)$ in the support set H when using the vector of attributes f. Therefore, the optimal vector of formal attributes f^* should satisfy the condition

$$Y_z(\Delta^*|f^*) = \min_{f \in B} Y_z(\Delta^*|f). \tag{5}$$

On the other hand, for chosen criterion of signal classification, the function $\lambda(f)$ directly characterizes the efficiency of the classification algorithm within the support set H. Therefore, it is desirable that the following condition holds:

$$f^* = \arg\min_{f \in B} U(\lambda(f)), \tag{6}$$

where the function $U()$ is positive, continuous and monotonically decreasing in the interval $[0, 100]$. Taking into account the fact that $\lambda \in [0, 100]$, the function $U(x) = 1/|1 + x|$ can be used as the function $U(\ )$. Recall that the vector f^* must obey the inequality

$$T(f^*) \leq G. \tag{7}$$

In the general case, the vector f^* that simultaneously satisfies conditions (5) - (7) may not exist. However, a suboptimal solution of the problem of choosing formal attributes is always possible. In this case, we solve the problem of searching for a vector f^* that renders minimum to the functional $Q(f) = (R\{Y_z\{\Delta^*|f\}\} + R\{U(\lambda(f))\})$ over the set B and ensures the validity of the inequality $T(f^*) \leq G$. Here, $R(x)$ is the scale function of the parameter x. This vector may be defined as follows:

$$f^* = \arg\min_{f \in B} W(f|G).$$

Here, $W(f|G) = E(G, f)Q(f)$ and

$$E(G, f) = \begin{cases} \infty, & \text{for } T(f) > G \\ 0, & \text{for } T(f) \leq G. \end{cases}$$

6. EXAMPLES OF SPACES OF FORMAL ATTRIBUTES FOR AUTOMATIC CLASSIFICATION

Consider several variants of special transformations over realizations of noiselike finite signals that allow one to use the results of these transformations as formal attributes in automatic classification of initial signals.

6.1. The Space of Structural Characteristics

Every finite signal s detected is characterized by the moment t_0 of its beginning and the moment t_F of its end. Let for some $\varepsilon > 0$, the variable r denote the maximal number of nonintersecting subintervals of duration ε that entirely belong to the interval $[t_0, t_F]$. Hereafter, realizations of a signal that correspond to these subintervals will be called fragments of the signal. We denote this set of intervals by $L(s)$. For every $l \in L(s)$, consider an energy characteristic of this signal, e.g.,

$$E(l) = \sum_{t \in l} s^2(t)/\varepsilon.$$

Here, $s(t)$ is the discrete count of the signal s that corresponds to the moment $t \in [t_0, t_F]$. By $A(l, \tau)$, we denote the function whose value is the number of intersections produced by the observed realization of the signal s and a constant level τ in the interval $l \in L(s)$. Let the set of observed images of signals I_E play the role of a training sample. Every signal s from S_E is naturally related to the set of pairs

$$J(s) = \{(E(l), A(l, \tau)|l \in L(s)\}.$$

Moreover, let

$$J(S_E) = \bigcup_{s \in S_E} J(s).$$

Using any one of the known methods of clustering [2], we partition the set $J(S_E)$ into n clusters. The number n is determined in an optimal way on the basis of intrinsic tendencies of clusterization that manifest themselves in attempts to partition the set $J(S_E)$ into different numbers of clusters. The methods of determining the number n are thoroughly described in [2]. Furthermore, every cluster is assigned a symbol of an alphabet (e.g., a letter in a Latin alphabet). In this case, syntactical description of a signal will be based on representing the realization of this signal in the form of a sequence of its fragments, each of which belongs to one of n clusters. Therefore, to define a syntactical description of a signal, it is sufficient to define the sequence of clusters that include the corresponding fragments of the signal. Thus, we uniquely determine the sequence of symbols of the chosen alphabet; this eventually provides a syntactical description of this signal.

Definition 1.

Let ξ and ε be two sequences of characters that consist of symbols of the alphabet A. For ξ and ε we introduce the following three types of admissible transformations over the set of sequences of symbols from A:

– substitution,

$$\xi a \varepsilon \longrightarrow [T(s, a, b)] \longrightarrow \xi b \varepsilon, \ a \neq b;$$

– exclusion,

$$\xi a \varepsilon \longrightarrow \{T(d, a)\} \longrightarrow \xi \varepsilon;$$

– inclusion,

$$\xi \varepsilon \longrightarrow [T(i, a)] \longrightarrow \xi a \varepsilon.$$

For any pair of sequences ξ and θ, we define the path J from ξ to θ as follows:

$$J(\xi \longrightarrow \theta) = T(1) \cdot T(2) \cdot \ldots \cdot T(n),$$
$$T(i) \in \{T(s, \cdot), T(d, \cdot), T(i, \cdot)\}. \tag{8}$$

The number $\text{LEN}[J(\xi \longrightarrow \theta)] = n$ of admissible transformations that occur in composition (8) is called the length of the path.

Definition 2.

The Levenshtein distance $D(\xi, \theta)$ is defined as the least number of transformations, whereby the sequence θ can be obtained from the sequence ξ [3].

Each transformation described above is put in correspondence with a positive function $W(\cdot)$, called the weight of transformation in the sequel. Obviously, the operations of exclusion and inclusion have greater weights. For these operations, the Euclidean distance between the origin of coordinates and the center of the cluster, where the character considered is excluded (included), can be taken as the weight function. It is expedient to use the distance between the centers of the clusters that correspond to the interchanged characters

as the weight function for the operation of subs[...] The weighted path length is defined as the sum[...] weights of constituent admissible transformatio[...]

$$\text{LEN} [*, J(\xi \longrightarrow \theta)] = \sum_{i=1}^{n} W(i).$$

Here, $W(i)$ is the weight of the ith transform[...] composition (8).

Definition 3.

The number

$$D(*, \xi, \theta) = \min_{J(\xi \to \theta)} \text{LEN} [*, J(\xi \longrightarrow \theta)]$$

is called the weighted Levenshtein distance.

To determine the distance $D(*, \cdot)$, one can [...] method of dynamic programming.

Thus, it is admissible to use the distance [...] whose metric properties are trivially tested, as [...] in the space of structural characteristics.

6.2. Subspace of Spectral Characteristic[...]

Let $F(n, t) \in R^m$ be an estimate of the spec[...] the signal power constructed by $2n$ points of m[...] ments starting from the tth point. For noiselike [...] it is quite expedient to assume that the element[...] belong to the space of n-dimensional vectors [...] Kendall metric $d_k()$, which is, by definition, rob[...] respect to random fluctuations of components[...] vector $F(n, t)$. Recall that the distance betw[...] n-dimensional vectors $F_1(n, t)$ and $F_2(n, t)$ i[...] mined in this space in the following way:

$$d_k(F_1(n, t), F_2(n, t)) = 1 - 2/(n(n - 1)) \sum_{l < k} \Delta[...]$$

$$\Delta_{lk}^j = \begin{cases} 1 & \langle F_j(n, t) \rangle_l < \langle F_j(n, t) \rangle_k \\ -1, & \langle F_j(n, t) \rangle_l > \langle F_j(n, t) \rangle_k \\ 0, & \langle F_j(n, t) \rangle_l = \langle F_j(n, t) \rangle_k. \end{cases}$$

The use of the sequence of vectors $F(n, t_1)$, [...] ..., $F(n, t_l)$, $t_l \in [t_0, t_f]$, which we denote by $\mathbf{F}(\mathbf{T},$ [...] $[t_0, t_f]$, as a formal attribute is promising for fir[...] nals. This sequence characterizes fairly w[...] dynamics of frequency–temporal properties of t[...] sified signal.

In this case, the function

$$\rho_k(F_1(\mathbf{T}, n), F_2(\mathbf{T}, n)) = \sum_{t_i \in T} d_k(F_1(n, t_i), F_2([...]$$

can be used as a metric.

The properties of the metric $\rho_k(\cdot)$ are easily [...]

7. RESULTS OF PRACTICAL APPLICATION

The described method of recognizing stochastic signals implemented in the form of a subsystem of automatic classification of seismic noise now successfully functions as a part of the system for automatic security of underground works at the Komsomol'skii mine (Noril'sk industrial region). At the stage of adaptation of the recognition subsystem the samples of signals that represent the most probable classes of noise that enters the input of the subsystem of signal registration accompanying the functioning of open-pit mining were introduced.

The following classes of signals were considered *a priori*:

(a) technological noise, including the noise of drilling equipment, punchers, pneumatic drills, blasting, ore chutes, and drilling;

(b) noises of relaxation processes, including local cracking of rocks and mining shocks.

Technological noise was represented in the training sample by 20 realizations of signals of each type. Relaxation processes were represented by 25 realizations of rock cracking and 10 realizations of mining shocks. As a result, the following statistic of the quality of classification was registered during the period from October 10, 1991 to March 30, 1992:

(1) technological explosions: 100% correct classification, the size of the test sample was 700 signals;

(2) ore chute: 95% correct classification; in the other 5% of the cases, ore chutes were misclassified as punching; the size of the test sample was 460 signals;

(3) punching: 90% correct classification; in the other 10% of the cases, punching was misclassified as ore chute; the size of the test sample was 600 signals;

(4) drilling: 100% correct classification; the size of the test sample was 156 signals;

(5) rock cracking: 92% correct classification; the other 8% of the samples were misclassified as technological explosions; the size of the test sample was 1560 signals;

(6) mining shocks: from October 10, 1991 to March 30, 1992 only one mining shock took place; it was classified correctly.

The test results were obtained according to the following scheme.

(1) The next automatically detected seismic signal was entered into the input of the classification system, where it was automatically classified and the moment of signal generation was fixed.

(2) Space coordinates of the source of the signal were computed simultaneously with automatic classification of the signal. Note that mining works are always carried out in accordance with daily regulations worked out in advance. Based on the computed coordinates of the signal source and the generation time of the signal and using the schedule of technological works, we can easily determine the technological type of works carried out in a given spatial region of mining at a given time. Next, we only need to compare the results of automatic classification with the results obtained from the spatial–temporal schedule of technological works. In addition, experts–operators were involved in the procedure of controlling the correctness of the results of classification. These experts analyzed visual images of previously classified signals and made conclusions concerning the correctness of classification. This was especially urgent for the classification of signals from relaxation events (mining shocks and rock cracking).

Thus, the statistic presented shows fairly high practical efficiency of the technique proposed for classifying stochastic broadband signals in the case of classifying seismic signals of various nature.

CONCLUSION

Application of the algorithm for the classification of stochastic signals proposed in the paper demonstrated high efficiency. This conclusion is supported by the statistics presented in this paper. One of the main advantages of this algorithm consists in the naturalness and simplicity of the practical implementation of the algorithm.

REFERENCES

1. Katkovnik, V.Ya., *Neparametricheskie Identifikatsiya i Sglazhivanie Dannykh* (Nonparametric Identification and Smoothing of Data), Moscow: Nauka, 1985.

2. Mandel', I.D., *Klasternyi Analiz* (Cluster Analysis), Moscow: Finansy i Statistika, 1988.

3. Levenshtein, V.I., Binary Codes Using Exclusions, Inclusions and Substitutions of Symbols, *Dokl. Akad. Nauk SSSR*, 1965, vol. 163, no. 4, pp. 845 - 848.