

Theoretical Aspects of the Patterns Recognition Statistical Theory Used for Developing the Diagnosis Algorithms for Complicated Technical Systems

A A Obozov¹, I N Serpik², G S Mihalchenko³ and G A Fedyaeva⁴

¹ Department of Thermal Engines, Bryansk State Technical University, Bryansk, Russia

² Department of Mechanics, Bryansk State Engineering and Technology University, Bryansk, Russia

³ Department of The Rolling Stock of Railways, Bryansk State Technical University, Bryansk, Russia

⁴ Department of Electronic, Radio-Electronic and Electro-Technical Systems, Bryansk State Technical University, Bryansk, Russia

Abstract. In the article, the problem of application of the pattern recognition (a relatively young area of engineering cybernetics) for analysis of complicated technical systems is examined. It is shown that the application of a statistical approach for hard distinguishable situations could be the most effective. The different recognition algorithms are based on Bayes approach, which estimates posteriori probabilities of a certain event and an assumed error. Application of the statistical approach to pattern recognition is possible for solving the problem of technical diagnosis complicated systems and particularly big powered marine diesel engines.

1. Introduction

Complicated technical systems (CTS) might be determined as the systems with the significant number of incoming, outgoing and internal parameters, their functional connectivity and also with a hardly viewed bloc structure [1]. Diversity and complexity of the processes following the running CTS, impacts of external factor on their functioning process create considerable difficulty of developing affective diagnostics algorithms.

At present, the greatest number of algorithms, proposed for CTS diagnosis, are presented in the form of so-called «functions faults matrices». As the examples we may see the algorithms developed for diagnosing marine diesel engines, which fully comply with the concept of “a complex technical system” [2,3,4]. When processing data with the use of such algorithms, the determined diagnostic decision is accepted. The advantage of such algorithms is, above all, their simplicity and logical clearness of the diagnostics procedure performed, that could be done in a manual way without involving any electronic and computer means.

There is a qualitatively different approach to solving the problem of constructing STS diagnosis algorithms based on the concepts of new sections of cybernetics - the statistical theory of pattern recognition and a selflearning systems theory. The subject matter of the latter is the range of problems connected with the objects recognition (classification) independently of their physical nature and developing technical means that are able to do such recognition. In case of using the *learning concepts*,



the recognition system (the recognition system could be a computerized diagnostics system for CTS) is not only assigned to have a recognition function on the basis of having the «ready for use» algorithm, but also to have the function of obtaining original teaching data regarding the analyzed object and finding the optimal pattern recognition algorithm. Such an approach would solve one of the major problems of implementing STD CTS – the problem of immediate obtaining the diagnosis algorithm in the short field tests. In this case, training is reduced to STD tests conducted consistently by CTS during its working condition and artificially administered by different defects (adjustment disorders). As a result of study based on the analysis of information coming from the sensors installed by STS, STD will form the diagnostic algorithm itself, which is subsequently used for fault recognition. Forming the problem it is necessary to build and lay in STD a rule according to which it will analyze incoming information and build fault diagnosis algorithm. Such a rule can be called a diagnosing system learning algorithm.

It should be emphasized that the positive effect of applying the patterns recognition of the statistical theory for solving a CTS technical diagnosis problem (particularly – for big marine diesel engines) has already been pointed out in literature [5, 6]. However, the practical realization of such approach has not been widely used. This is probably explained by the lack of clear engineering techniques. It is important to reiterate the essence of statistical pattern recognition methods, to show their feasibility and benefits, using specific examples relating to the diagnosis of STS.

2. The methodical aspects of pattern recognition statistic theory

Let us make some preliminary clarifications regarding methodology of the statistical theory of the pattern recognition [7, 8, 9].

The object to be recognized can be characterized by the set of parameters x_1, x_2, \dots, x_n , n – the total number of parameters. If the geometrical interpretation is used, the set of parameters generates n -dimensional vector x (vector of observation) within the *parameters space* (the parameter space can also be called *space of observation*). The objects with similar characteristics are unified in *classes*. In parameter space, each class forms compact gathering vectors, which takes the space specific for the class. In other words, when using a probabilistic-statistical approach, we may say, that the definite statistical distribution of vector x will correspond to each class within parameter space.

In figure 1, there is a simple two-dimensional example of vector x statistical distribution, corresponding to two recognizable classes (for example, good and bad conditions of CTS). If these two distributions are well-known, there is a possibility of fixing boundary $D(x_1, x_2)=0$ in the parameter space which will divide the space into two regions R_1 and R_2 . Now, depending on the direction in which area R_1 or R_2 gets realization of the recognizable object observation vector (depending on what sign function $D(x_1, x_2)$ takes), we are able to make a decision about belonging of the x -vector to any object. In other words, we may come to the conclusion «what object we are currently monitoring».

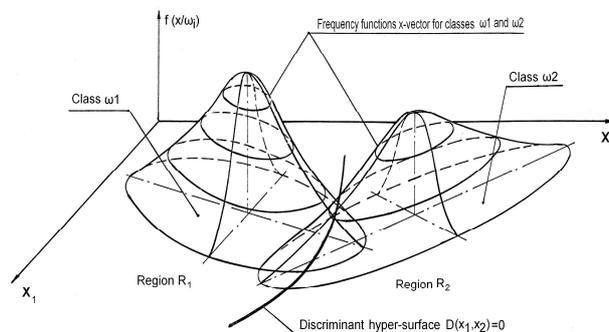


Figure 1. Space of parameters divided into areas R_1 and R_2 , used when solving the pattern recognition problem.

Function $D(x_1, x_2)$ is called a *discriminant* function, the recognition method that uses the discriminant function is called *discriminant analysis*. When doing recognition of the statistical objects

(namely these problems are to be solved when fixing CTS diagnostics) there are the cases when the classes intersect in the parameters space. In other words, the classes partially lay over each other. In this case, it is not possible to make the classes recognition absolutely correct and it is important to minimize this error.

The key conception of the statistical theory of patter recognition (and similar scientific disciplines: statistical decision theory, discriminant analysis, etc.) is a predictive error (a probability of predictive errors in decision-making). The level of predictive errors depends on quality of the recognition system and the algorithm used in the system. In addition, the level of an error depending on the property of classes are to be recognized, that is the level of an error depends on the class dividing degree within the space of parameters.

It is widely known that the probability of the predictive error (in case of observing dichotomous classification) is made up of error probabilities of the first and secondary type.

Let ω_1 and ω_2 be the classes to be recognized.

A probability of *the first type error* (object erroneously refers to class ω_2 , although it actually belongs to class ω_1) is calculated as

$$P_{(2/\omega_1)} = P(\omega_1) \int_{R_2} f(x/\omega_1) dx, \tag{1}$$

where $P_{(2/\omega_1)}$ – the probability of the first type error; $P(\omega_1)$ – the probability of $x \in \omega_1$; $f(x/\omega_1)$ – the conditional frequency of distributing the observation x -vector, corresponding to class ω_1 ; R_2 – the x -space area for integration.

A probability of *the secondary type error* (the object erroneously refers to class ω_1 , although it actually belongs to a class ω_2) is calculated as

$$P_{(1/\omega_2)} = P(\omega_2) \int_{R_1} f(x/\omega_2) dx, \tag{2}$$

where $P_{(1/\omega_2)}$ – the probability of the secondary type error; $P(\omega_2)$ – the probability of $x \in \omega_2$; $f(x/\omega_2)$ – conditional frequency of distributing the observation x -vector, corresponding to class ω_2 ; R_1 – the x -space region for integration.

The total probability level of predictive error P_{error} is calculated as

$$P_{\text{ou}} = P_{(2/\omega_1)} + P_{(1/\omega_2)}.$$

All the material mentioned above is illustrated in Figure 2.

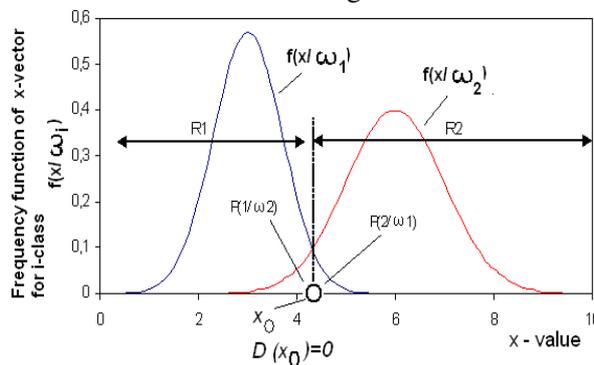


Figure 2. Definition of the probabilities for the first and secondary type errors when recognizing two classes (one-dimensional example, x_0 – a boundary for taking decisions).

The effective statistical recognition methods are the methods based on Bayes theorem. These methods also admit the possibility of calculating the probabilities of the x -vector belonging to any classes to be recognized.

According to Bayes theorem a posterior probability of the x -vector belonging to class ω_i is

$$P(\omega_i / x) = \frac{P(\omega_i) f(x / \omega_i)}{f(x)}, \quad (3)$$

where $P(\omega_i)$ – the probability of $x \in \omega_i$; $f(x/\omega_i)$ – conditional frequency of distributing the observation \mathbf{x} -vector, corresponding to class ω_i ; $f(\mathbf{x})$ – weighted frequency of distributing the possibility of the \mathbf{x} -vector over all classes.

$$[f(x) = \sum_{i=1}^k P(\omega_i) f(x / \omega_i), k - \text{the number of all classes to be recognized.}$$

Let us return to an overview of the example of only two classes of recognition (dichotomous classification). If the decisive rule recognizes the \mathbf{x} -vector according to a provided maximum for belonging probability, that is,

$$P(\omega_1 / x) > P(\omega_2 / x) \rightarrow x \in \omega_1,$$

$$\text{and otherwise: } P(\omega_1 / x) < P(\omega_2 / x) \rightarrow x \in \omega_2 \quad (4)$$

Taking into account (3), rule (4) might be rewritten as:

$$f(x / \omega_1) P(\omega_1) > f(x / \omega_2) P(\omega_2) \rightarrow x \in \omega_1,$$

$$\text{and otherwise } x \in \omega_2. \quad (5)$$

Formula (5) might be rewritten as:

$$l(x) = \frac{f(x / \omega_1)}{f(x / \omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \rightarrow x \in \omega_1,$$

$$\text{and otherwise } x \in \omega_2. \quad (6)$$

Function $l(\mathbf{x})$ is called a *likelihood function*, and value $P(\omega_2) / P(\omega_1)$ is called a *boundary value* of the likelihood function for the given decisive rule.

The practical experience shows that the majority of recognition procedures might be brought to objects (classes) recognition tasks with the parameter frequency function following the multidimensional normal distribution law. In such cases, the conditional frequency of \mathbf{x} -vector distribution might be written as:

$$f(x / \omega_i) = \frac{1}{(2\pi)^{\frac{1}{2}k} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right], \quad (7)$$

where k - dimension of the \mathbf{x} -vector; μ - vector of expected values \mathbf{x} of distributing the ω_i class parameters; Σ - covariance matrix (components of the \mathbf{x} -vector); $|\Sigma|$ - determinant of the covariance matrix; t - symbol of the matrix (vector) transpose; $-I$ - symbol of matrix inversion.

Likelihood function (6) for the examples of the normal distribution law after logographic transformation looks like:

$$\frac{1}{2} (x - \mu_1)^t \Sigma_1^{-1} (x - \mu_1) - \frac{1}{2} (x - \mu_2)^t \Sigma_2^{-1} (x - \mu_2) + \frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} > (<) \ln \frac{P(\omega_2)}{P(\omega_1)} \Rightarrow x \in \omega_1 (x \in \omega_2), \quad (8)$$

where μ_1 and μ_2 are vectors of expected values \mathbf{x} for classes ω_1 and ω_2 ; Σ_1 and Σ_2 - covariance matrix parameters for classes ω_1 and ω_2 ; $|\Sigma_1|$ and $|\Sigma_2|$ - determinants of covariance matrixes for classes ω_1 and ω_2 .

Taking into account (8), we might make the following conclusion: if the left part (8) exceeds the logarithm of a class priory probability fraction, then the x -vector belongs to class ω_1 , otherwise, the x -vector belongs to class ω_2 .

For the examples when the matrixes are equal (or «close» to each other) $\Sigma_1 = \Sigma_2 = \Sigma$, the given decisive rule looks significantly simpler:

$$x^t \Sigma^{-1} (\mu_1 - \mu_2) - \frac{1}{2} (\mu_1 + \mu_2)^t \Sigma^{-1} (\mu_1 - \mu_2) > (<) \ln \frac{P(\omega_2)}{P(\omega_1)} \Rightarrow x \in \omega_1 (x \in \omega_2) \quad (9)$$

The first term of inequality $x^t \Sigma^{-1} (\mu_1 - \mu_2)$ - is the line combination, which consists of the x -vector components. Value $-\frac{1}{2} (\mu_1 + \mu_2)^t \Sigma^{-1} (\mu_1 - \mu_2)$ is an absolute term of the discriminant function. It might be strengthened or weakened depending on value $\ln \frac{P(\omega_2)}{P(\omega_1)}$. The latter means some

shifting in a parallel way of the hyper surface, that divides classes within the parameter space. If one admits priory probabilities for classes equal to $(P(\omega_1) = P(\omega_2) = 0,5)$, then the rule is transformed into the rule of *maximum likelihood*

$$x^t \Sigma^{-1} (\mu_1 - \mu_2) - \frac{1}{2} (\mu_1 + \mu_2)^t \Sigma^{-1} (\mu_1 - \mu_2) > (<) 0 \Rightarrow x \in \omega_1 (x \in \omega_2) \quad (10)$$

It should be noted that if there is information on the apriori probability, values of the CTS grade are diagnosed by correcting a free member of the discriminant function. It is easy to move from the maximum likelihood rule to the rule of the minimum probability of erroneous decisions.

Statistical theory for patterns recognition also refers to a number of decision rules optimal in a determined sense (a method of minimal risk, a minimax method, Neumann-Pearson's method and others) which are widely depicted in special literature.

We have considered the basic methodological principles discussed in the work which are used for recognizing the statistical objects. Given principles make it possible to develop the effective recognition algorithms (particularly – recognizing the different technical condition of CTS and making learning systems of technical diagnostics) which are based on Bayes theorem.

In particular, there was a rule of *maximum likelihood* which was viewed applicable for solving the task of dichotomous classification (recognizing the projects belonging to two classes). In cases of necessity of recognizing several classes, the task might easily be brought to consecutive dichotomy. Especially for this purpose, it is necessary to build discriminant functions for every couple of classes and to form a matrix.

$$\| D_{ij}(x_{ij}) \| = \begin{pmatrix} D_{11}(x_{11}), & D_{12}(x_{12}), & \dots, & D_{1k}(x_{1k}) \\ D_{21}(x_{21}), & D_{22}(x_{22}), & \dots, & D_{2k}(x_{2k}) \\ \dots & \dots & \dots & \dots \\ D_{k1}(x_{k1}), & D_{k2}(x_{k2}), & \dots, & D_{kk}(x_{kk}) \end{pmatrix} \quad (11)$$

It is obvious, that every discriminant function included in matrix (11) could enclose its own x -vector qualitatively different from others, because the description of different recognizable classes could be done with different parameters. The calculated values of discriminant functions that are located symmetrically to the matrix main diagonal (for any x -vector realization), will always be equal (in absolute value) but with an opposite sign. Discriminant functions located on the matrix main diagonal are insensible (serve for recognizing one class relatively the same class). According to decisive rule (9), the algorithm for recognizing k classes (the classes of CTS diagnosis) is the search of such a matrix row (of matrix (11)) which has the positive signs of all elements of the row (after x -vector substitution), while searching the matrix diagonal elements is ignored. Only one matrix row could be positive all round. It means this row will point to the recognized class.

Examining the discriminant matrix changes in different times $\| D_{ij}(x_{ij}) \|$ we can create a *time derivative matrix*

$$\| d[D_{ij}(x_{ij})] / dt \| .$$

A further extrapolation method points the time when the \mathbf{x} -vector goes from «current» class ω_i into «new» class ω_j . The analysis of the formed *time intervals matrix* can approximately be calculated as:

$$\tau(\omega_i \rightarrow \omega_j) = -\frac{D_{ij}(x_{ij})}{d[D_{ij}(x_{ij})] / dt}, \quad (12)$$

(where $D_{ij}(x_{ij}) > 0$, because current position of the \mathbf{x} -vector is located in ω_i and it gives the possibility to determine the moment and type of the most probable crossing of marine diesel engines diagnostics.

3. Conclusion

The article viewed the decisive rules based on «classic» Bayes scheme of decision-making. This approach is able to create effective recognition algorithms used especially for classifying hard recognizable classes of statistic nature. As it was shown, algorithms based on Bayes scheme make it possible to identify the objects, give probabilistic assessment of the identification error and solve a problem of predicting the system behavior. Methods of statistical theory of pattern recognition might be applied for analyzing the condition of complicated technical systems (CTS) and particularly for analyzing the technical condition of big marine engines. The number of publications confirm the fruitfulness of using the viewed theory for solving the problem of marine diesel technical diagnostics [10-14].

References

- [1] Shibanov G P 1977 *Functioning Control of big systems* 356
- [2] Nikitin E A and Ctanslavskiy L V 1987 *Diesel diagnosis* 224
- [3] Shishkin V A 1986 *Fault Analysis and Fault Avoiding in Diesels of Ships* 192
- [4] Mozgalevskiy A V and Gaskarov D V 1975 *Technical diagnostics* 205
- [5] Birger I A 1978 *Technical diagnostics* 240
- [6] Pavlov B V 1966 *Cybernetics method of technical diagnostics* 150
- [7] Fukunaga Keinosuke 1972 *Introductions to statistical pattern recognition School of electrical Engineering, Academic Press* 367
- [8] Edward A 1972 *Patrick Fundamentals of Pattern Recognition, Department of Electrical Engineering Purdue University* 407
- [9] Gorelik A G and V A 1989 *Skripkin Recognition Methods* p 232
- [10] Levin M I and Obozov A A 1986 *Application of selflearnig method for solving the problem for diagnosing the marine diesels* **9** 41-46
- [11] Moek E, Shtrikert X 1986 *Technical diagnostics of marine machinery and mechanisms* p 232
- [12] Bley D C et al. 1987 *American Nuclear Society*
- [13] Bley D C 1991 *Trans Amer Nucl Soc, PL&G* **64** 443-444
- [14] Uemukai Teruhito, Nomura Tomohiro and Higashino Ichiro 1984 *Bulletin of the ISME* **27(231)** 1957-1964