RESEARCH ARTICLE OPEN ACCESS

# Studying the influence of training sample volume on the average risk of technical diagnostics

I.M. Korzhov*, R.P. Mygushchenko*, P.V. Shchapov**, O.Yu. Kropachek***
** Department of Information and Measuring Technologies and Systems, National Technical University "Kharkov Polytechnic Institute", Kharkiv, Ukraine*
*** Department of Industrial and Biomedical Electronics, National Technical University "Kharkov Polytechnic Institute", Kharkiv, Ukraine*
**** Department of Theoretical Foundations of Electrical Engineering, National Technical University "Kharkov Polytechnic Institute", Kharkiv, Ukraine*
*Corresponding Author: I.M. Korzhov*

**ABSTRACT**
The article deals with the influence of training sample volume on the average risk of technical diagnostics. The object of diagnostics is an industrial extruder. The measurement signals used for diagnostics are the heat signals of the heating zones and the vibration signals of the bearing unit in the extruder. The first stage of the measurement signals processing is to obtain informative auto-coherence indices. The second stage is to make a logical decision on the control object status. Training samples corresponding to a priori known technical states of the control object are used to form a mathematical model of the second stage of processing. The volume of such training samples has a significant impact on the average risk (reliability) of control and diagnostics. The article shows that for a real control object (industrial extruder) there is an opportunity to optimize the training sample volume by minimizing the average risk of diagnosis.

***Keywords* -** diagnostics, control, auto-coherence, reliability, statistical risk

Date Of Submission: 24-01-2019 Date Of Acceptance:08-02-2019

## I. INTRODUCTION

The problem of improving information systems effectiveness for control and diagnostics of the technical state of industrial objects with dynamic properties is inextricably linked with an increase in volumes of measurement information that characterizes typical variants of dynamic violations [1, 2].

Studying information systems designed for control and diagnostics of the dynamic disruption states of industrial objects is an evaluation task regarding limited volume of measurements and parameters characterizing equipment dynamics during test changes in the technical state [2].

The article studies the influence of the training sample volume on the average risk of technical diagnostics on the example of an industrial extruder using auto-coherence indices as informative parameters.

## II. SYNTHESIS OF MATHEMATICAL MODEL OF MEDIUM RISK

Consider, in general, the discrimination model in the form of a linear function with dependent features. Covariance matrices Σ for auto-coherence indices $X_1$, …, $X_4$ are known [1, 3], as there is always the possibility for increasing the number of multiple measurements $N_a$, and matrices Σ are calculated using the volume $N_a$. The size of the training sample $N_b$ in the diagnosed states must satisfy the simple condition: $N_b \geq 3$ [2, 3].

If the matrix Σ is known and the decision making rule is invariant to linear transformations, it is convenient to consider the following: if the matrix Σ is single $E_p$ ($p$ is the number of informative features) [3, 4]. Such a transformation allows using linear Fisher discriminator for analysis. In this case, the vector $\bar{x}$ of auto-coherence indices is random:

$$\begin{cases} \bar{x} \approx NORM\left(\mu_{(0)}, E_p\right), \text{ if } S \in S_0; \\ \bar{x} \approx NORM\left(\mu_{(1)}, E_p\right), \text{ if } S \in S_0. \end{cases}$$

Such a transformation of the matrix Σ into the matrix $E_p$ rearranges the Mahalanobis distance [3]:

$$\delta_M = \left(\mu_{(1)} - \mu_{(0)}\right)' \Sigma^{-1} \left(\mu_{(1)} - \mu_{(0)}\right),$$

where the prime means the transformation operation in the geometric distance between the states $S_1$ and $S_2$ in the $p$-dimensional space of the vector $\bar{x}$:

$$\delta_g = \left(\mu_{(1)} - \mu_{(0)}\right)' \left(\mu_{(1)} - \mu_{(0)}\right)$$

In this case, the average risk estimation $P^*$ is a random variable $\varphi$:

$$\varphi = \sum_{k=1}^{2} P_k \cdot \Phi\left\{ -\frac{\delta_g\left(\xi_1 + \delta_g\sqrt{N/2}\right)}{2\sqrt{\left(\xi_1 + \delta_g\sqrt{N/2}\right)^2 - \xi_3}} + (-1)^k \xi_2 \sqrt{N/2} \right\}, \quad (1)$$

where $\Phi\{\ \}$ is the probability integral; $\xi_1$, $\xi_2$, $\xi_3$ are independent random variables, the first two of which have a standard normal distribution $\xi_1 \approx$ NORM(0,1) and $\xi_2 \approx$ NORM(0,1), and the third is the central X-squared distribution with $p$ degrees of freedom $\xi_2 \approx \chi^2_p$ [3, 4].

The mathematical expectation $\overline{\varphi}$ of the random variable (1) is to be found, given that the mathematical expectations for $\xi_1$, $\xi_2$, $\xi_3$ correspondingly equal $\overline{\xi}_1 = \overline{\xi}_2 = 0$ and $\overline{\xi}_3 = p$. Consider also that $\overline{\varphi}$ is the function $\overline{\xi}_1, \overline{\xi}_2, \overline{\xi}_3$, and $\Phi(-z) = 1-\Phi(z)$, where $z$ is the argument for probability integral [3, 4]. Then $\overline{\varphi}$, as the average risk, is determined by the expression:

$$\overline{\varphi} = 1 - \Phi\left\{ \frac{\delta^2}{2}\left(\delta^2 + \frac{2p}{N}\right)^{-\frac{1}{2}} \right\} \quad (2)$$

Expression (2) for the mathematical expectation considering the average risk of diagnostics points to the important properties of the discrimination procedure with small volumes $N$ of training samples:

1. Increase of $N$ reduces the average risk, as the argument $z$ of the integral $\Phi\{\ \}$ goes up;
2. Increasing the average risk leads to an increase in the number of $p$ informative features (auto-coherence indices $X_1, \ldots, X_p$), as $z$ decreases;
3. If the geometric distance $\delta_g$ between the diagnosed states is constant, the relation $2p/N = \kappa$ must be taken into account.

Any reduction of N must be compensated by a decrease in the number of p signs (otherwise κ increases and, accordingly, the average risk increases (the argument z of the probability integral decreases)).

## III. STUDYING THE EFFECTS OF AVERAGE RISK MINIMIZATION FOR HEAT SIGNALS

Equation (2) was used to study the problem of forming informative features system, based on auto-coherence indices $X_1 - X_4$, which were calculated for heat signals of heating zones in the industrial extruder. As informative properties of these indices are not the same, they were ranked by reducing geometric distance $\delta_g$, and further consecutive development of the informative features system with increasing space dimension. This formation was carried out separately for a pair of states $S_1 - S_2$ [1, 3]. The following equation (3) was used to estimate the geometric distance $\delta_g$:

$$\delta_g = \frac{|M_1 - M_2|}{\sqrt{\left(\sigma_1^2 + \sigma_2^2\right)/2}} \quad (3)$$

The following equation was used to form the space dimension $n$

$$\delta_n = \sqrt{\sum_{k=1}^{n} \delta_k^2}, \quad (4)$$

where $\delta_n$ is the geometric distance between the states of the studied pair in the space dimension $n$.

While ranking, the $X$ value was changed to encoded $Y$ with appropriate changes in the indexes. Table 1 shows the geometric distances separately for each of the indices $X_1 - X_4$ $\delta_g$.

**Table 1.** Geometric distances $\delta_g$

| Initial value | Coded values | $\delta_g(S_1 - S_2)$ |
|---|---|---|
| $X_3$ | $Y_1$ | 0.579831085 |
| $X_2$ | $Y_2$ | 0.482922299 |
| $X_4$ | $Y_3$ | 0.401651309 |
| $X_1$ | $Y_4$ | 0.255228700 |

Table 2 shows the geometric distances $\delta_n$ between states in the increasing space dimension $n$ for a pair of states. The minimum risk value $\overline{\varphi}$ is highlighted.

**Table 2.** Geometric distances $\delta_n$ between states

| $n$ | $\delta_n$ | $\overline{\varphi}$ |
|---|---|---|
| 1 | 0.579831 | 0.794883 |
| 2 | 0.754598 | 0.747907 |
| 3 | 0.854834 | 0.729374 |
| 4 | 0.892123 | 0.736441 |

Table 2 clearly shows the existence of a minimum risk for the space dimension of features, which is less than the maximum dimension $n = 4$ [1, 3]. Graphic illustration of the existence of such a minimum is presented in Fig. 1.

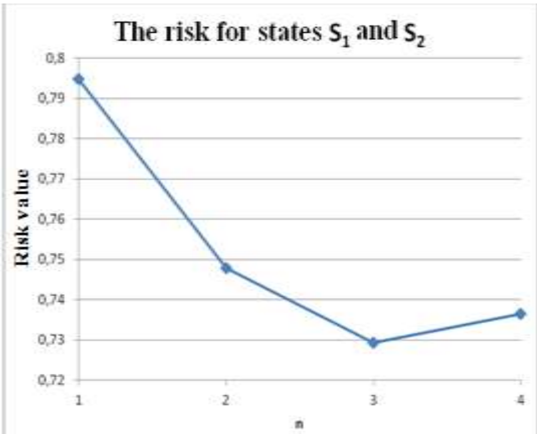


**Fig 1.** The risk for a pair of states $S_1 - S_2$

## IV. STUDYING THE EFFECTS OF AVERAGE RISK MINIMIZATION FOR VIBRATION SIGNALS

The equation (2) was used to study the problem of forming informative features system, based on auto-coherence indices $X_1 - X_4$ [1, 3], which were calculated for the vibration signals of the bearing unit in the industrial extruder. As informative properties of these indices are not the same, they were ranked by reducing geometric distance $\delta_g$, and further consecutive development of the informative features system with increasing space dimension. This formation was carried out separately for the pairs of states $S_1 - S_2$, $S_2 - S_3$, $S_1 - S_3$. The following equation (3) was used to estimate the geometric distance $\delta_g$.

The following equation was used to form the space dimension *n* (4).

While ranking, the *X* value was changed to encoded *Y* with appropriate changes in the indexes. Table 3 shows the geometric distances separately for each of the indices $X_1 - X_4$ $\delta_g$.

**Table 3.** Geometric distances $\delta_g$

| Initial value | Coded values | $\delta_g$ $(S_1 - S_2)$ | $\delta_g$ $(S_2 - S_3)$ | $\delta_g$ $(S_1 - S_3)$ |
|---|---|---|---|---|
| $X_1$ | $Y_1$ | 0.603911 | 3.231845 | 2.973713 |
| $X_3$ | $Y_2$ | 0.408747 | 3.182352 | 2.676992 |
| $X_4$ | $Y_3$ | 0.178144 | 1.209599 | 0.465311 |
| $X_2$ | $Y_4$ | 0.046090 | 0.361331 | 0.448790 |

Table 4 shows the geometric distances $\delta_n$ between states in the increasing space dimension *n* for a pair of states. The minimum risk value $\overline{\varphi}$ is highlighted.

**Table 4.** Geometric distances $\delta_n$ between states

| *n* | $(S_1 - S_2)$ | | $(S_2 - S_3)$ | | $(S_1 - S_3)$ | |
|---|---|---|---|---|---|---|
| | $\delta_n$ | $\overline{\varphi}$ | $\delta_n$ | $\overline{\varphi}$ | $\delta_n$ | $\overline{\varphi}$ |
| 1 | 0.6039 | 0.7681 | 3.2318 | 0.0249 | 2.9737 | 0.0397 |
| 2 | 0.7292 | 0.7445 | 4.5357 | 0.0017 | 4.0012 | 0.0058 |
| 3 | 0.7507 | 0.7641 | 4.6942 | 0.0012 | 4.0281 | 0.0060 |
| 4 | 0.7521 | 0.7858 | 4.7081 | 0.0013 | 4.0530 | 0.0062 |

Table 4 clearly shows the existence of a minimum risk for the space dimension of features, which is less than the maximum dimension n = 4 [1, 3] regardless of the number of studied states. Graphic illustration of the existence of such a minimum is presented in Fig. 2.

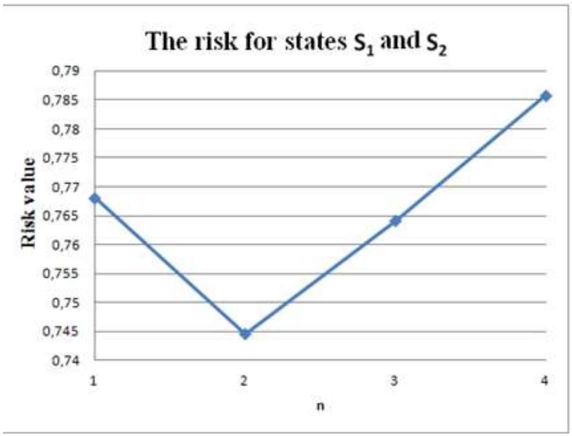


**Fig 2.** The risk for a pair of states $S_1 - S_2$

## V. CONCLUSIONS

Tables 2 and 4, figures 1 and 2 indicate the possibility of features space optimization depending on the *N* volume of the training sample and the geometric distance between the diagnosed states.

## REFERENCES

[1]. Shhapov P.F., Mygushhenko R.P., Kropachek O.Ju.,Korzhov I.M.Doslidzhennyakorelyatsiynykh modeley spektral'noyi nestatsionarnosti vypadkovykh sygnaliv [Research correlation models of spectral non-stationary random signals]. Metrologiya ta prylady.Naukovo-vyrobnychyy zhurnal [Metrology and Instruments. Scientific and production magazine]. Kharkiv, 2018, no. 5 (73), pp. 11–14.

[2]. Shhapov P.F., Avrunin O.G. Povyshenie dostovernosti kontrolya i diagnostiki ob" ektov v usloviyakh neopredelyennosti : monografiya [Improving the reliability of monitoring and diagnostics of objects in conditionsofuncertainty:monograph].Kharkov, KnNADU Publ., 2011. 191 p.

[3]. Korzhov I.M., Kropachek O.Ju. Realizatsiya obchyslyuval'nykh protsedur systemnogo informatsiynogoperetvorennyanestatsionarnykhvibrosygnalivzasobamylabview[Implementation of computational procedures for system information transformation of non-stationary vibration signals by means of labview]. Sistemy upravlinnya, navigatsiyi ta zv'yazku. Zbirnyk naukovykh prats'. Poltava, PNTU Publ., 2017, vol. 6 (46), pp. 76–80.

[4]. Basvil' M, Vilski A, [et al.] Obnaruzhenie izmenenija svojstv signalov i dinamicheskih system [Detection of changes in signal properties and dynamic systems]. Moscow, Mir Publ., 1989. 278 p.

I.M. Korzhov" Studying the influence of training sample volume on the average risk of technical diagnostics" International Journal of Engineering Research and Applications (IJERA), vol. 9, no.2, 2019, pp 64-66