ISSN 2076-2429 (print)
ISSN 2223-3814 (on line)

UDC 681.5

G.Y. Shcherbakova, Ph. D., ass. prof.,
V.N. Krylov, Dr. techn. sciences, prof.,
R.A. Pisarenko, master,
Odesa nat. polytechnic univ.

INFORMATION TECHNOLOGY OF PARAMETERS PREDICTION WITH ADAPTIVE CLUSTERING IN THE SPACE OF THE WAVELET TRANSFORM

Г.Ю. Щербакова, В.М. Крилов, Р.О. Писаренко. **Інформаційна технологія прогнозування параметрів за допомогою адаптивної кластеризації у просторі вейвлет-перетворення.** Запропоновані інформаційна технологія прогнозування за допомогою адаптивної кластеризації у просторі вейвлет-перетворення і підсистема автоматизованого прогнозуванні параметрів партії виробів електронної техніки. Ця інформаційна технологія дозволяє скоротити тривалість виробничих випробувань виробів у разі значної зміни їх у часі, високому рівні перешкод і малих вибірок даних.
Ключові слова: адаптивна кластеризація, гіперболічне вейвлет-перетворення, шум.

Г.Ю. Щербакова, В.Н. Крылов, Р.А. Писаренко. **Информационная технология прогнозирования параметров с помощью адаптивной кластеризации в пространстве вейвлет-преобразования.** Предложены информационная технология прогнозирования с помощью адаптивной кластеризации в пространстве вейвлет-преобразования и подсистема автоматизированного прогнозировании параметров партии изделий электронной техники. Эта информационная технология позволяет сократить длительность производственных испытаний изделий в случае значительного изменения их во времени, высоком уровне помех и малых выборок данных.
Ключевые слова: адаптивная кластеризация, гиперболическое вейвлет-преобразование.

G.Y. Shcherbakova, V.N. Krylov, R.A. Pisarenko. **Information technology of parameters prediction with adaptive clustering in the space of the wavelet transform.** Information technology of prediction using adaptive clustering in the space of the wavelet transform and subsystem of automated parameters prediction of the electronic products parties were suggested. This information technology can shorten manufacturing test of products in case of its significant change over time, a high noise level and small samples of data.
Keywords: adaptive clustering, hyperbolic wavelet transform.

An implementation high quality requirements electronic product (EP) requires significant investments of time and resources for manufacturing tests. Quality assurance leads to greatly time and cost requirements for manufacturing testing. This leads to increases of development terms, manufacturing application and overstates the cost of products. One of the ways to provide the quality of electronic devices is to reduce the duration of manufacturing tests. There are two basic ways to reduce the duration of manufacturing tests — accelerated testing and automated prediction of technical parameters of objects [1]. The first way requires significant cost. Therefore, for further studies in the work second direction is chosen - automated prediction of products parameter changes in time and estimation fault probability. Automated prediction may be based on two approaches. The first approach allows us to extrapolate parameter value using one-dimensional models, if the EP has determining or generic parameter, or in the case of multi-dimensional non-correlated parameters. If the parameters of EP are correlated, this approach is not applicable. The second approach — statistical classification — is used when the conditions of EP set are described by multidimensional functions. EP is divided into classes according to the degree of working capacity at given instant of time in the future or according to the product life variety.

Disadvantage of the approach — calculation of covariance matrices of the tested EE set up to the failure. This approach requires a large amount of priori researches. Due to the significant working life of EP it isn't practicable to receive such training data.

Statistical classification is used if it is sufficient to indicate the class of this EP. It consists of two stages — clustering and classification [1]. During clustering data is divided into clusters on the basis of compactness, to optimize the quality functional. This functional can be multiextremal due to the complex shape of clusters. Clustering methods, based on the evaluation of the gradient, do not provide sufficient noise immunity, and methods which estimate subgradient have high error. Therefore, for clustering in these conditions adaptive multistart method in the space of the wavelet transform (WT) was developed. To reduce the impact of the above deficiencies, this method of clustering is used in this work.

Objective of this work is to develop information technology of parameters prediction to shorten manufacturing tests with help of noise proof method of adaptive multistart clustering in the space of the WT. To achieve the goal following problems are solved: analysis of the main models of EP parameters scattering during process of exploitation, development of a method for predicting EP parameters to reduce the duration of manufacturing tests using multistart clustering method in the space of the WT and the procedure of implementing this forecasting method; investigation of prediction quality, information technology and adaptive predicting subsystem development.

The process of manufacturing EP is affected by many destabilizing random factors that determine the change in the EP parameters in time. Among them are breakdowns in process due to the tool and equipment wear, irregularity in stuff and parts delivery, changes in technical requirements, incorrect product yield planning etc. Variation in time of EP parameters is also caused by peculiarities of the measurement process.

Processes of change in the output EP parameters at the initial period of usage in most cases lead to one of the following types of scattering. Thus, for oven-controlled quartz oscillators of standard signals, generators and reference-voltage sources, high-stable voltage divisors and others, the initial period of operation is characterized by a slight change of expectations and almost linear time variation of parameters standard deviation (SD). If a change of output parameter is caused by discharge of measuring system power supply, permanent change of some external factor, such as, for example, ambient temperature or pressure, the stray field of this parameter is characterized by systematic drift of a center of scatter. The width of the stray field increases monotonically in time due to the increase of magnitude of standard deviation. There are cases when the output parameters of the stray field of measuring devices are characterized over time by one or more of the extremums of their expectation. This can occur during exploitation or storage of such devices under conditions with significant changes of influencing factors. Cluster with such distribution of parameters for a few time steps can be shifted so much, that the parameters of some objects will move to the next cluster. This variant of dispersion model will be considered in the work.

During parameters estimation based on the statistical classification at fixed points of time in the group of EP compact subgroups (clusters) with the same properties are allocated. For example, due to the fact that EP are collected from parts of different manufacturers, such clusters may have different degrees of monotony in parameters dependence on operating time during tests. By applying an adaptive clustering in the proposed method noise immunity is improved and estimation error of the cluster parameters drift is reduced, since the initial cluster centers parameters for subsequent points in time are evaluated from the analysis of the previous moment [1]. With this estimation class number to which each object will belong at next instant of time is determined. In this work we suggest to make prediction using Markov chains [5]. The elements of fuzzy classification are used in forecast to determine the transition probabilities of the EP states. This takes account of the distances of all elements to the centers of clusters. Suggested forecasting method uses the statistical classification with self-learning, which includes adaptive clustering and classification itself. Clustering methods are used for clustering and determining transition probabilities of the EP states, and the statistical classification methods — for decision-making when predicting their parameters [1].

ISSN 2076-2429 (print)
ISSN 2223-3814 (on line)

In the approach to parameters estimation used in this paper we don't predict the values of certain EP parameters in time, but only the class, which each object will belong. Due to this the forecast can be automated on the basis of the proposed in [1] adaptive clustering method in the space of the WT. This method has low error, and allows clustering in multidimensional feature space with high noise level and the small volume of the investigated samples. Therefore, this method is proposed to use in predicting of the EP parameters to shorten manufacturing testing and reducing forecast error in the application of Bayesian procedures for Markov chains.

The problem of parameters forecasting is as follows. Suppose there is *N* objects (EP), each of which is characterized by a set of *k* parameters. To estimate the behavior of these objects at time their parameters are measured at specified intervals. In the *k*-dimensional space *X* object *j* at the point of time *t* is determined by the point $x_j(t)=(x_j^1(t), x_j^2(t),\ldots, x_j^k(t))$. The set of points in the parameter space $x_j(t_1),\ldots, x_j(t_n)$ corresponds to the sequence of time $t_1, t_2,\ldots, t_n$ of parameters measurement and is a part of the object trajectory. Vector $x_j(t)=(x_j^1(t), x_j^2(t),\ldots, x_j^k(t))$ describes the state of *j*-th unit at time *t*. This means that the relative position of a points set $x_1(t),\ldots, x_N(t)$ in *k*-dimensional parameter space reflects the actual classification (classification according to the parameters in time) of the objects. To identify this structure, multistart clustering method is used. By applying it at the moment t_1 points in space *X* are divided into *r* classes (clusters). The number of clusters is selected by taking into account the local density of the data in feature space based on the hypothesis of the λ -compactness and subsequent search of extremum of the target function formed with the λ -compactness and then remains constant [2…4].

Such an approach has been adopted because the research is conducted with small samples of noisy data, as well as with large difference between the number of elements in the clusters. The large difference between the number of elements in the clusters can occur when the product of one party are separated by parameters, for example, two clusters (in classes of accuracy, reliability parameters, etc.). In such case, not only the distance between the elements are encouraged to consider, but also the relationship between them, that is to apply the methods to determine the number of clusters based on the hypothesis of the λ-compactness. This hypothesis takes into account the normalized distance between the elements of the cluster and the local density of ensemble in their surroundings [2, 3]. The indicator of the partition quality is the value describing the average change of λ-distance at step [4]:

$$K_i=\frac{f(i)}{f(i-1)},\quad i\in[1,n-1], \tag{1}$$

where *f(i)* — the average λ-distance for all clusters for the current version of the separation.

Number of clusters is determined by the maximum of *f*(*i*), which is sought by multistart optimization method in the space of the WT based on the methodology described in [1], due to multimodality of the target function. Since the standard deviation and mathematical expectation of the parameters change over time, which changes the boundaries of clusters up to the intersection, the right choice of their number and increasing noise immunity and reduction uncertainty of clustering procedures in the first control point substantially determines the reliability of the forecast generally [3]. The block diagram of the parameters prediction system is shown in Fig. 1. It operates by the following information technology [1, 5].

1. The first stage involves the allocation of informative parameters set and their measurement.

2. At the instant of time t_1 *n* points in space *X* are separated on *r* classes (clusters) using multistart clustering method. The concept of the class standard $a_i(t)$, $i=1,\ldots,r$ (usually the center of the class) is introduced [5]. For the objects the distances $R_{ij}(t)$, $i=1,\ldots, r$, $j=1,\ldots,$ is calculated.

3. At time t_1 the transition probabilities $p_{ji}^{(1)}=p_{ji}(t_1)$ are calculated as follows [5]:

$$p_{ji}^{(1)} = \frac{\alpha_j^{(1)}}{R_{ji}^{(1)}}, \qquad (2)$$

where $\alpha_j^{(1)} = \left(\sum_{l=1}^{r} \frac{1}{R_{jl}^{(1)}} \right)^{-1}$ — normalization factor.

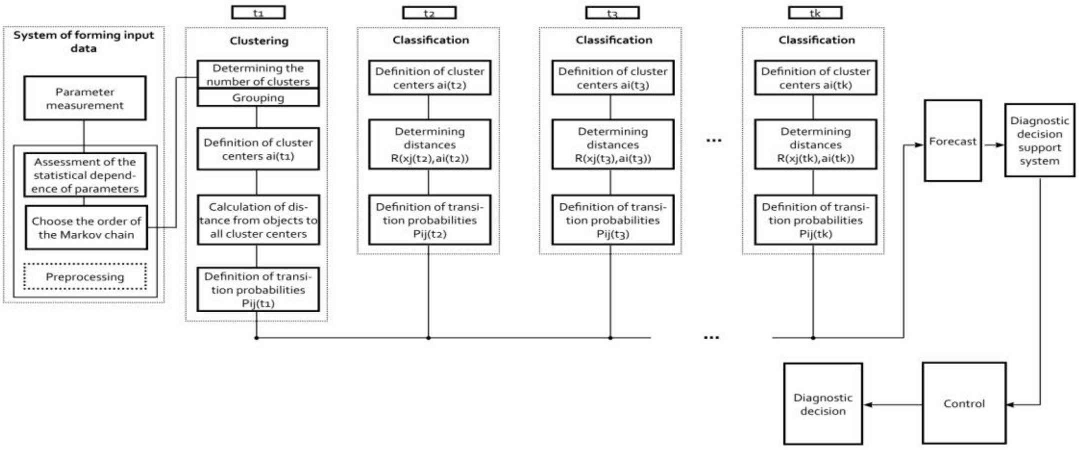


Fig. 1. Adaptive predicting system flowchart

4. At time t_2, each point $x_j(t_2)$ with the classification based on the multistart optimization method [1] is related to one or another class, obtained in the first step. After that values of cluster centers (standards) $a_i(t_2)$, $i=1,\ldots,r$ and the distances $R(x_j(t_2),\ a_i(t_2))$ from points $x_j(t_2)$ to new standards are recalculated. This procedure is performed for all m time points.

5. At the s-th step transition probability matrix (2) is modified by the procedures [5]:

— if the point coincides with the standard of a class, it is considered that the probability for this point to stay in the current class is one, and the probability of transition to a different class is 0;

— if the j-th point doesn't coincide with the standard of the class, transition probabilities are calculated as:

$$p_{ji}^{(s)} = \gamma \left[p_{ji}^{(s-1)} + \left(\frac{1 + sign(\Delta R_{ji}^{(s)})}{2} - p_{ji}^{(s-1)} sign(\Delta R_{ji}^{(s)}) \right) \Delta \tilde{R}_{ji}^{(s)} \right], \qquad (3)$$

where γ — a scale factor, determined by the normalization of the transition probabilities $\sum_{i=1}^{r} p_{ji}^{s} = 1$:

$$\gamma = \frac{1}{1 + \left(\frac{1 + sign(\Delta R_{ji}^{(s)})}{2} - p_{ji}^{(s-1)} sign(\Delta R_{ji}^{(s)}) \right) \Delta \tilde{R}_{ji}^{(s)}}, \qquad (4)$$

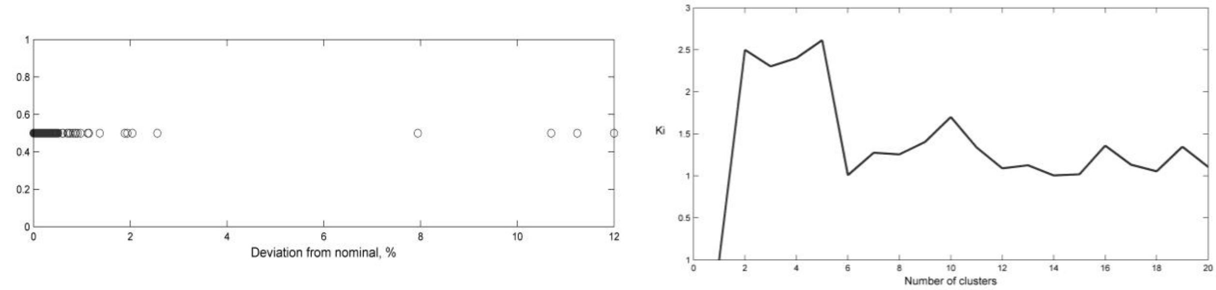
$$sign(z) = \begin{cases} 1, & if\ z \ge 0, \\ -1, & if\ z < 0 \end{cases}; \quad \Delta R_{ji}^{(s)} = R_{ji}^{(s-1)} - R_{ji}^{(s)}; \quad \Delta \tilde{R}_{ji}^{(s)} = \frac{R_{ji}^{(s-1)} - R_{ji}^{(s)}}{R_{ji}^{(s-1)} + R_{ji}^{(s)}}.$$

6. Forecast of object belonging is realized according to the constructed transition matrix. For this it is common to use the Bayesian scheme, according to which the object belongs to that class, where $p_{ji0} = \max_{i=1,\ldots r} p_{ji}$. In case of equality of transition probabilities p_{ji} object belongs to the class with the lowest number [5].

ISSN 2076-2429 (print)
ISSN 2223-3814 (on line)

The proposed information technology was tested on the example of parameters prediction of metallodielectric resistors. These resistors were divided into subgroups according to the noise level from -26 dB to +19 dB. At 24, 168, 1,000, 5,000 and 10,000 hours resistance value measurements were carried out. Noise level and deviation of resistance from their nominal were used as predictors. Based on the data from studies [1] (the mathematical expectation and standard deviation from the nominal value of the resistor in the subgroups) the sample parameters of 410 resistors were synthesized.

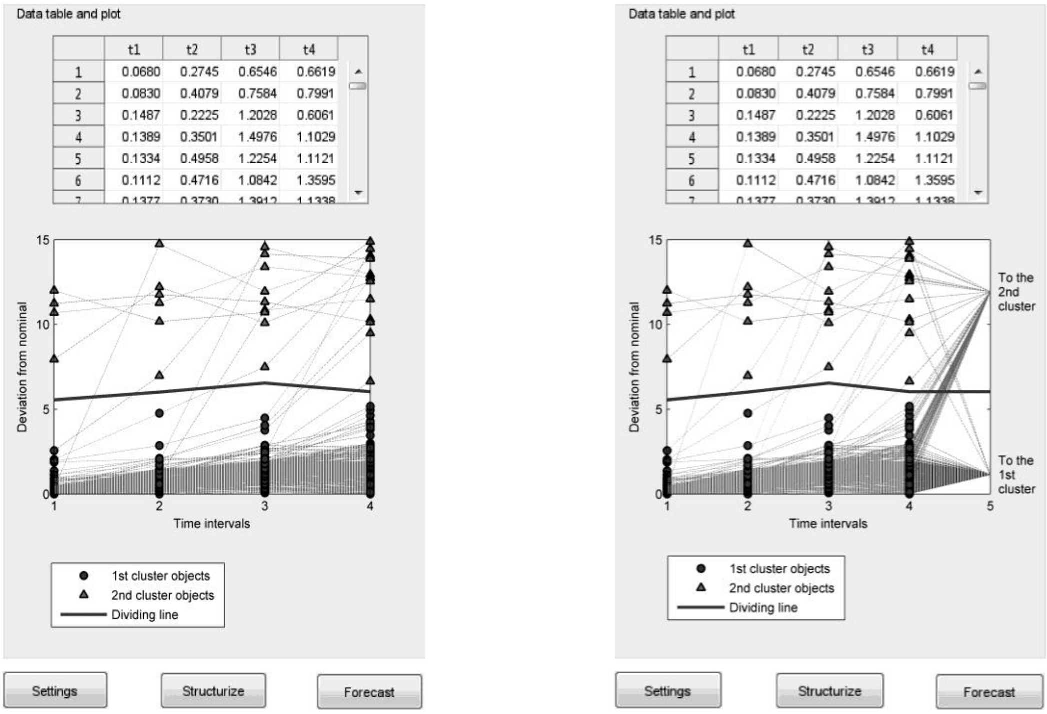
At the first point of time these data [1] (Fig. 2, a) were separated by an adaptive clustering into 2 classes. The number of classes was chosen on the basis of high quality index (Fig. 2, b).



a *b*

Fig. 2. The results of accelerated testing (a) and the quality index for number of clusters definition (b)

Fig. 3, a shows the trajectory of parameters in four time sections. Further, it was performed forecasting of objects belonging to clusters at time t_5 (10,000 hours) and evaluation of the prediction quality.



a *b*

Fig. 3. The results of classification in four time sections (a) and prediction of objects belonging to clusters at time t_5 (b)

The result is shown in Fig. 3, b. Estimation of the prediction reliability was calculated:

$$P = \frac{n}{N}, \tag{5}$$

where n — the number of objects with correctly predicted class number; N — the number of objects.

For the investigated data was obtained P=0,7, what meets the requirements of practice. When forecasting using gradient algorithms in phases 2 and 4, P is lower in 1,4 times, this suggests a decrease of the average risk in the application of the proposed method.

Thus, the proposed method of parameters prediction reduced the time of manufacturing test for investigated group of resistors in 2 times. This reduction in testing time is determined by the non-uniform intervals in the measurement of the studied sample (the forecast is carried out for one step in time, from 1 to 4 and between 4 and 5 step — by 5000 hours). This result allows us to recommend the developed method and information technology of its implementation for selection of the EP, which are intended for use in long-running equipment for continuous working high-duty equipment applications.

Literature

1. Щербакова, Г.Ю. Прогнозирование параметров с помощью адаптивной кластеризации / Г.Ю. Щербакова // Радіоелектронні і комп’ютерні системи. — 2010. — № 5 (46). — С. 129 —134.
2. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. — Новосибирск: Изд-во ин-та математики, 1999. — 270 с.
3. Щербакова, Г.Ю. Определение количества кластеров при прогнозировании состояния электронной аппаратуры / Г.Ю. Щербакова, В Н. Крылов, С.Г. Антощук // Электроника и связь. — 2010. — № 3 (56). — С. 91 — 95.
4. Крисилов, В.А. Решение задачи таксономии на основе гипотезы компактности при анализе данных / В.А. Крисилов, Н.О. Блудян, С.А. Юдин // Искусственный интеллект. — 2005 — № 4. — С. 699 — 707.
5. Дорофеюк, Ю.А. Методы адаптивного структурного прогнозирования / Ю.А. Дорофеюк // Информационные технологии и вычислительные системы. — 2010. — № 3. — С. 34 — 39.

References

1. Shcherbakova, G.Y. Prognozirovanie parametrov s pomoshh'ju adaptivnoj klasterizacii [Parameters prediction using adaptive clustering] / G.Y. Shcherbakova // Radioelektronni i komp’juterni systemy [Radioelectronics and computer systems]. — 2010. — # 5 (46). — pp. 129 — 134.
2. Zagoruiko, N.G. Prikladnye metody analiza dannyh i znanij [Applied methods of data analysis and knowledge] / N.G. Zagoruiko. — Novosibirsk: Izdatel'stvo instituta matematiki [Publishing House of the Institute of Mathematics], 1999. — 270 p.
3. Shcherbakova, G.Y. Opredelenie kolichestva klasterov pri prognozirovanii sostojanija jelektronnoj apparatury [Determination of the number of clusters in state predicting of the electronic equipment] / G.Y. Shcherbakova, V.N. Krylov, S. G. Antoshchuk // Elektronika i svjaz' [Electronics and Communication]. — 2010. — # 3 (56). — pp. 91 — 95.
4. Krisilov, V. A. Reshenie zadachi taksonomii na osnove gipotezy kompaktnosti pri analize dannyh [Solution of the problem of taxonomy based on the hypothesis of compactness in the data analysis] / V.A. Krisilov, N.O. Bludyan, S.A. Yudin // Iskusstvennyj intellekt [Artificial Intelligence]. — 2005 — # 4. — pp. 699 — 707.
5. Dorofeyuk, Y. A. Metody adaptivnogo strukturnogo prognozirovanija [Adaptive structural forecasting methods] / Y.A. Dorofeyuk // Informacionnye tehnologii i vychislitel'nye sistemy [Information technologies and computer systems]. — 2010. — # 3. — pp. 34 — 39.

Reviewer Dr. techn. sciences, prof. Odesa nat. polytechnic univ. Kostenko V.L.

Received December 20, 2012.