Predictive Associative Search Models in Variable Structure Control Systems

NATALIA N. BAKHTADZE, EKATERINA A. SAKRUTINA,
ELENA F. JHARKO
Laboratory of Identification of Control Systems
V.A. Trapeznikov Institute of Control Sciences
65, Profsoyuznaya, Moscow, 117997
RUSSIA
sung7@yandex.ru

Abstract: - Predictive associative search models found a wide application in advanced control systems, decision making systems, play a fundamental role in any activity concerned with the signal processing process. In the paper, *associative search* identification models in variable structure control systems are considered. The application of this approach is especially effective whilst compensating for insufficient lab data for model development. In such a case, fuzzy specification of certain process variables using process knowledge base is practiced. A methodology of the synthesis of a variable structure control system is presented, as well issues of the stability of a model built by use of the associative search are considered, in the aspect of the spectrum analysis of the multi-scale wavelet expansion.

Key-Words: - system identification, control system, information model, virtual model, associative search, fuzzy model, knowledge base, wavelets

1 Introduction

Variable structure control systems (VSCS) are a class of control systems, providing an effective possibility to solve main problems of the control theory – stabilization problem and tracking problem – by use of automatically switched algorithms.

VSCS provide a rational system performance in accordance to a set dynamics without constructing an adaptive system model (structure and parametric one). In paper [1], the control invariance with respect to parametric disturbances was proven.

The methodology of the synthesis of variable structure systems is based on introducing a fundamental notion of new kind feed-backs (operator and coordinate-operator ones). The approach is forming an operator (an algorithm of forming a control implemented by a controller) transforming signals-coordinates (functions in the time) as an element of a set of stabilizing feed-backs. A parameterization of the set enables one to impose a one-to-one relationship between the signals-coordinates and signals-operators.

In paper [1], a constructive approach to the design of such algorithms is presented, enabling the sliding mode, for second order linear systems. However, the VSCS ideology provides a possibility to design control systems of such a kind under the conditions of the parametric uncertainty not for linear dynamic plants only. In the present paper, a possibility of applying the principle of the operator feed-backs for non-linear as well as time-varying systems is demonstrated.

To solve such a problem, one assumes under forming the control algorithm at each time instant (what corresponds to the definition of a variable structure system) applying information on the system status/state (current and archive one), in other words: application of all its dynamic previous history. Meanwhile, not a conventional control system with a feed-back identifier is formed: a system with operator feed-back is formed, in which the signal-operator is formed by use of a virtual plant model created at each time instant on the basis of intelligent analysis of persistently updated date on the plant dynamics.

2 Variable structure systems using inductive knowledge

Let us consider a scheme of a control system with additional dynamic error feed-back, displayed in Fig. 1.

In systems investigated in [1], the operator $\mathbf{S}(t)$ produced the sign change. As to the general case, in this block some, generically, non-linear transformation of the signal s is implemented:

$$r(t) = \mathbf{S}(s(t), \sigma(t)) \tag{1}$$

Here **S** is a non-linear operator that is formed on the basis of analysis a specific control problem. In the most general case, **S** may be, for instance, logic production operator implementing both conventional and fuzzy control.

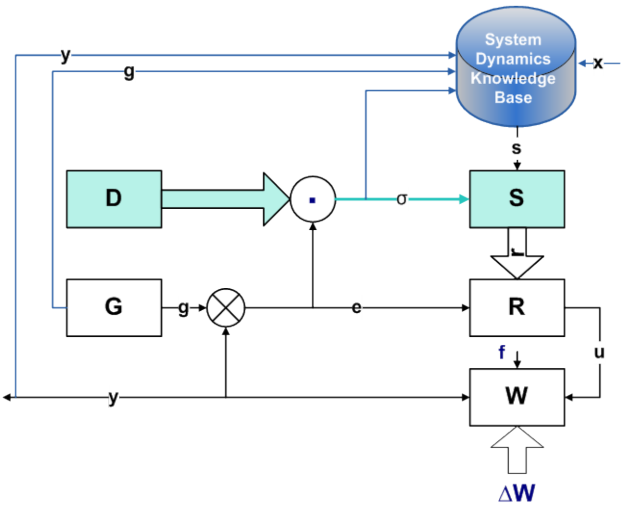

Fig. 1. Scheme of control system with additional dynamic error feed-back

To form at each time instant t the signal-operator r producing varying the structure of the regulator **R**, besides the dynamic error $\sigma(t)$ one should use additional a priori information on the system. In this connection, let us recollect that the unique information code sets the system state, its information "portrait". How can one use this information (interpreted as the signal $s(t)$), and how is $s(t)$ formed?

$s(t)$ is to provide to **S** an intelligent (based on knowledge, that is regularities, formed and updated at each time instant on the basis of the history data analysis) prediction information model of the plant. This model provides to **S** an "information support" to make a decision by the system on changing the controller structure. Based on this support, the signal-operator $r(t)$ is formed.

Taking into account the fact that the same value of the dynamic error $\sigma(t)$ may take place under different sets of signal values (involving operator ones), we come to the inference that the signal $s(t)$ may be represented as a regression (in the general case, nonlinear) of the following variables:

$$s(t) = \mathbf{F}[(g(t), g(t-1), \dots, g(t_0), \\ \sigma(t-1), \dots, \sigma(t_0), y(t-1), \dots, y(t_0), \\ u(t-1), \dots, u(t_0), t]. \tag{2}$$

A principal distinction of the intelligent model from a conventional predicting model is the following. The intelligent (that is based on knowledge revealed from the data analysis) model is, in its entity, a situational ("virtual") one, that is accounting particularities of the current state, hence, at each time instant this model may have a new structure. This its attribute enables one to synthesize VSCS for non-linear, and, involving, time-varuing systems.

The operator **F** (also, in the general case, being non-linear) is to be formed by use an approach based on the data analysis. The history system data analysis will enable one to reveal certain regularities (inductive knowledge). An example of such knowledge may consist in relating the system input (a vector in corresponding vector space) at each current time instant t to a certain domain of this space (clustering).

To determine $s(t)$, in particular, one may apply the technique of the *associative search* [2], being a search method based on the analysis of the previous history of state dynamics of a plant under study and constructing *virtual models*.

3 The associative search method

Algorithms based on knowledge revealed from history system data (inductive knowledge, persistently enriched) implement an intelligent approach to constructing identification models. The intelligence is applying knowledge (*Knowledge Based*) revealed from history data on the basis of their analysis (*Data Mining*).

The process of knowledge processing in the intelligent system is reduced to recovering (associative search of) *knowledge* over its fragment. Meanwhile, the *knowledge* may be interpreted as associative connections between *images*. As an image, we will use "sets of indicators", that is components of input vectors, input variables.

The criterion of closeness between images may be formulated in very different manners. In the most general case, it may be represented as a logic function, the predicate. In a particular case, when sets of indicators are vectors in *n*-dimensional space, the criterion of closeness may be a distance in this space.

The associative search process may be implemented either as a process of recovering the image over partially given indicators (or recovering a knowledge fragment under the conditions of incomplete information; as a rule, just this process is simulated in different models of the associative memory), or as a process of searching another images that are associatively connected with the given one, attached to other time instants.

In papers [2-4], an approach to form the support on decision making on the control is proposed, based on dynamic modeling the associative search procedure. Results of adoption of the associative search algorithms developed by the authors for industrial processes of the chemical and petroleum manufacturing, processes of control in intelligent power networks (smart grids), trading processes, transport logistic processes.

The method of the *associative search* consists in constructing *virtual* predicting models. The method assumes constructing predicting model of a dynamic plant, being new under each t, by use of a set of history data ("associations") formed at the stage of learning and adaptively corrected in accordance to certain criteria, rather than approximating real process in the time.

Within the present context, linear dynamic model is of the form:

$$y_N = \sum_{i=1}^{m} a_i y_{N-i} + \sum_{j=1}^{r_s} \sum_{s=1}^{S} b_{j,s} x_{N-j,s}\,, \quad \forall j = \overline{1,N}, \tag{3}$$

where: y_N is the prediction of the output plant at the time instant N, x_N is the vector of input actions, m is the memory depth in the output, r_s is the memory depth in the input, S is the dimension of the input vectors, a_i and $b_{j,s}$ are the tuned coefficient, meanwhile $x_{N-j,s}$ are selected disregarding the order of the chronological decreasing, have been referred as the *associative pulse*.

Let us note that this model is not classical regression one: there are selected certain inputs in accordance to a certain criterion, rather than all chronological "tail".

The algorithm of deriving the virtual model consists in constructing at each time instant an approximating hypersurface of the space of input vectors and single-dimensional outputs. To construct the virtual model, corresponding to some time instant, from the archive there are selected input vectors being in a certain sense close to the current input vector. An example of selecting the vectors is described below. The dimension of this hypersurface is selected heuristically. Again, by use of classical (non-recursive) least squares (LS) method there is determined the output value (modeled signal) in the next time instant.

Meanwhile, each point of the global non-linear surface of the regression is formed in the result of using linear "local" models, in each new time instant.

In contrast to classical regression models, for each fixed time instant from the archive there are selected input vectors being close to the current input vector in the sense of a certain criterion (rather than the chronological sequence as it is done in regression models). Thus, in equation (3) r_s is the number of vectors from the archive (from the time instant 1 to the time instant N), selected in accordance to the associative search criterion. At each time segment $[N-1, N]$ there is selected a certain set of r_s vectors, $1 \le r_s \le N$. The criterion of selecting the input vectors from the archive to derive the virtual model in the given time instant over the current plant state may be as follows.

Let us introduce as a distance (a norm in $\Re^S$) between points of the S-dimensional space of inputs the value:

$$d_{N,N-j} = \sum_{s=1}^{S} \left| x_{N,s} - x_{N-j,s} \right|, \forall j = \overline{1,N}, \tag{4}$$

where $x_{N,s}$ are components of the input vector at the current time instant N.

By virtue of a property of the norm («the triangle inequality», we have:

$$d_{N,N-j} \le \sum_{s=1}^{S} \left| x_{N,s} \right| + \sum_{s=1}^{S} \left| x_{N-j,s} \right|, \quad \forall j = \overline{1,N}, \tag{5}$$

Let for the current input vector x_N:

$$\sum_{s=1}^{S} \left| x_{N,s} \right| = d_N. \tag{6}$$

To derive an approximating hypersurface for the vector x_N let us select from the archive of the input data such vectors x_{N-j}, $j = \overline{1,N}$ that for a set D_N the condition will be hold:

$$d_{N,N-j} \le d_N + \sum_{s=1}^{S} \left| x_{N-j,s} \right| \le D_N, \quad \forall j = \overline{1,N}, \tag{7}$$

where D_N may be selected, for instance, from the condition:

$$D_N \ge 2 d_N^{max} = 2 \max_j \sum_{s=1}^{S} \left| x_{N-j,s} \right|. \tag{8}$$

If in the selected domain there will be not enough quantity of inputs to apply the LS method, that is the corresponding system of linear equations will be unsolvable, then the selected criterion of selecting points in the space of inputs may be weakened due to increasing the threshold D_N.

Under the assumptions that the input actions meet the Gauss-Markov conditions, the estimates obtained via the LS method are unbiased and statistically effective.

4 Solving the system of linear equations for the LS method procedure

However, solving the problem of constructing virtual closed-loop models (as well as the conventional identification synthesis) are characterized by considerable difficulties. Under the closed-loop case, the control process is with depended values, and the question on the existence of identifying strategies is not trivial. Optimal controllers generate linear state feed-backs; this leads to an degenerate problem [5].

For the identification of dynamic plants by use of the associative search technique in the degenerate case, Moore-Penrose method [6, 7] is applicable to obtain pseudo-solutions to the system of linear equations under using the least squares procedure. Applying the method to solve the identification problem is presented in paper [8].

5 Clustering based associative search

In order to increase the computation capability at the stage of learning and under subsequent real plant performance, one of the data mining methods – clustering (dynamic classification, automated grupping data, unsupervised learning) is used. There are known numerous such methods: hierarchical algorithms, k-means algorithm, minimal covering tree algorithm, nearest neighbor method, etc. All they determine (in the dependence on the time, in contrast to the classification) the belonging of the point in the multi-dimensional space by one of the domain, which the space is partitioned on.

As a result, each investigated point in the multi-dimensional space may be related to a group by assigning a cluster label to it. In the problem of the associative search to select input vectors being "close" to the current one, the cluster label is determined in accordance to the associative pulse (the criterion of the associative selection), and to derive the virtual models the vectors are selected inside the corresponding cluster.

In the problem of determining the signal-operator, the method of the associative search enables one to predict $r(t)$ on the basis of data on belonging the point $s(t)$ to one of the clusters of the space

$$\mathbf{S}\{g(t), g(t-1), \dots, g(t_0), \\ \sigma(t), \dots, \sigma(t_0), y(t-1), \dots, y(t_0), \\ u(t-1), \dots, u(t_0), t\}. \quad (9)$$

Such an approach enables one to avoid the need of information on the structure of the control plant **W**. The plant is even not needed to be linear. The control quality will depend on the amount of analyzed and clustering criterion.

6 Simulation results

To derive virtual predicting model of the full power of a three-phase network, a model of the form was constructed:

$$S(t) = a_1 S(t-1) + b_1 Ia(t-1) + \\ +b_2 Ua(t-1), \quad (10)$$

where $S(t)$ is prediction of the full power at the next step, $S(t-1)$ is the available value of the power, $Ia(t-1)$ is the available value of the current of the phase A, $Ua(t-1)$ is available value of the voltage of the phase A.

In Fig. 2, the comparison of the real process with the prediction by use of the linear model and prediction by use of the associative search is displayed.

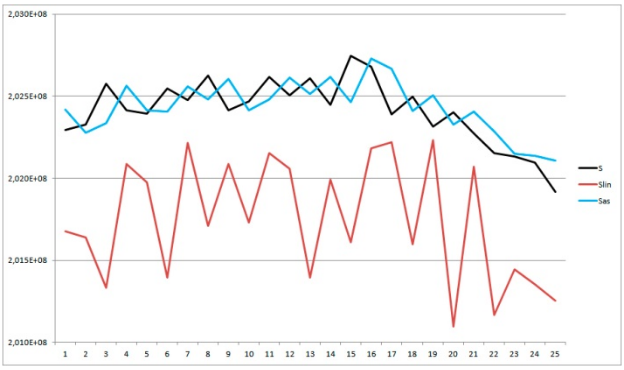

Fig. 2. The comparison of the real process (S) with the prediction by use of the linear model (Slin) and the prediction by use of the associative search (Sas)

7 Fuzzy virtual models

The application of fuzzy models for decision-making under fuzziness and uncertainty conditions is justified in the following cases:

- if one or more factors of the quality indexes dynamics are weakly or not formalized
- if the dynamics under investigation is described by sophisticated nonlinear relationships.

Production rules technique allowing to a certain extent to model human individual's thinking style is a key method of knowledge representation in modern systems employing expert knowledge. Any production rule consists of premises and a conclusion. Several premises in a rule are allowed; in such a case they are combined by logical operators AND, OR.

Fuzzy systems are based on production-type rules with linguistic variables used as premise and conclusion in the rule [11].

By renaming the variables, the linear dynamic plant's model (1) can be represented as follows:

$$Y_N = \sum_{i=1}^{n+m} a_i X_i.$$

We define a *fuzzy model* as a system with $n+m$ input variables $\mathbf{X} = \{X_1, X_2, \ldots, X_{n+m}\}$ defined over the input reasoning domain

$$DX = DX_1 \times DX_2 \times \ldots \times DX_n,$$

and a single output variable Y defined over the output reasoning domain DY. Crisp values of X_i and Y will be denoted by x_i and y respectively.

Fuzzy definitional domain of the i-the input variable X_i is denoted by $LX_i = \{LX_{i,1}, \ldots, LX_{i,l_i}\}$, where l_i is the number of linguistic terms on which the input variable is defined; LX_{ik} specifies the name of the k-th linguistic term. Similarly, $LY = \{LY_1, \ldots, LY_{ly}\}$ is the fuzzy definition domain of the output variable, l is the number of fuzzy values; LY_j is the name of the output linguistic term.

The rule base in the fuzzy *Mamdani system* is a set of fuzzy rules such as:

$$R_j: LX_{1,j_1}\text{AND}\ldots\text{AND}LX_{n,j_n} \to LY_j \tag{11}$$

The j-th fuzzy rule in the singleton-type system looks as follows:

$$R_j: LX_{1,j_1}\text{AND}\ldots\text{AND}LX_{n,j_n} \to r_j \tag{12}$$

where r_j is real number to estimate the output y.

The j-th rule in *Takagi–Sugeno model* (1985) looks as follows:

$$\begin{aligned} &R_j: LX_{1,j_1}\text{AND}\ldots\text{AND}LX_{n+m,j_{n+m}} \to \\ &r_{0j} + r_{1j}x_1 + r_{2j}x_2 + \ldots + r_{(n+m)j}x_{n+m} \end{aligned} \tag{13}$$

where the output y is estimated by a linear function. Thus, the fuzzy system performs the mapping $L: \Re^{n+m} \to \Re$.

The grade of crisp variable x_i membership in the fuzzy notion LX_{ij} is determined by membership functions $\mu_{LXij}(x_i)$. The *rule base* is determined by the criterion of minimum output error defined by one of the following expressions:

$$\frac{\sum_{i=1}^{K}|f(\boldsymbol{x}_i) - L(\boldsymbol{x}_i)|}{K}, \quad \frac{\sqrt{\sum_{i=1}^{K}\left(f(\boldsymbol{x}_i) - L(\boldsymbol{x}_i)\right)^2}}{K}, \quad \ldots \max_{i\in K}|f(\boldsymbol{x}_i) - L(\boldsymbol{x}_i)| \tag{14}$$

where K is the number of samples.

The choice of a fuzzy model depends on the plant's type and identification objective. For complex nonlinear dynamic plants, such as moving objects, where the accuracy requirements are predominant, the choice of Takagi-Sugeno [12] model looks reasonable. In the problems of knowledge formation from data (as linguistic rules) or the search of associative relations in a dataset, the Mamdani fuzzy system must be used. The *singleton-type* system may be used in both identification and knowledge formation tasks.

Singleton-type fuzzy model specified by rules (15) performs the mapping $: \Re^{n+m} \to \Re$, where the fuzzy conjunction operator is replaced by a product, and the operator of fuzzy rules aggregation, by summation. The mapping L is defined by the following expression:

$$L(\mathbf{x}) = \frac{\sum_{i=1}^{q} \mu_{LX\,1i}(x_1)\cdot\mu_{LX\,2i}(x_2)\cdot\ldots\cdot\mu_{LX\,(n+m)i}(x_{n+m})\cdot r_i}{\sum_{i=1}^{q} \mu_{LX\,1i}(x_1)\cdot\mu_{LX\,2i}(x_2)\cdot\ldots\cdot\mu_{LX\,(n+m)i}(x_{n+m})}, \tag{15}$$

where: $\mathbf{x} = [x_1, \ldots, x_{n+m}]^T \in \Re^{n+m}$; q is the number of rules in a fuzzy model; $n+m$ is the number of input variables in the model; $\mu_{LX_{ij}}$ is the membership function.

The expression for L mapping in a Takagi-Sugeno model looks as follows:

$$L(\mathbf{x}) = \frac{\sum_{i=1}^{q} \mu_{LX\,1i}(x_1)\cdot\ldots\cdot\mu_{LX\,(n+m)i}(x_{n+m})\cdot(r_{0i}+\ldots+r_{(n+m)i}x_{n+m})}{\sum_{i=1}^{q} \mu_{LX\,1i}(x_1)\cdot\mu_{LX\,2i}(x_2)\cdot\ldots\cdot\mu_{LX\,(n+m)i}(x_{n+m})}.$$

In Mamdani fuzzy systems, fuzzy logic techniques are used for describing the input vector's $\boldsymbol{x}$ mapping into the output value y, for example, Mamdani approximation or a method based on a formal logical proof.

Assume that one or more variables in (3) are fuzzy. In the real life, this may mean the fuzzification of weekly recommendation provided by major investment banks that in this case are considered as experts.

Generally, (3) can be represented as a fuzzy Takagi–Sugeno (TS) model.

The fuzzy TS model consists of a set of production rules with linear finite difference equations in the right-hand member (for simplicity, a single input case, i.e., P=1, is considered):

If

$y(t-1)$ is Y_1^{θ},…, $y(t-r)$ is Y_r^{θ}, $x(t)$ is X_0^{θ},…, $x(t-s)$ is X_r^{θ},

then

$$\begin{aligned} y^{\theta}(t) &= a_0^{\theta} + \textstyle\sum_{k=1}^{r} a_k^{\theta} y(t-k) + \\ &+ \textstyle\sum_{l=0}^{s} b_l^{\theta}\, x(t-l),\ \theta = 1, \ldots, n, \end{aligned} \tag{16}$$

where:

$$\mathrm{a}^{\theta} = (a_0^{\theta}, a_1^{\theta}, \ldots, a_r^{\theta}),$$
$$\mathrm{b}^{\theta} = (b_0^{\theta}, b_1^{\theta}, \ldots, b_s^{\theta})$$

are adjustable parameter vectors;

$\mathbf{y}(t-r) = (1, \mathrm{y}(t-1), \ldots, \mathrm{y}(t-r))$ is the state vector;

$\boldsymbol{x}(t-s) = (x(t), x(t-1),\ldots, x(t-s))$ is an input sequence;

$Y_1^\theta,\ldots,Y_r^\theta$, $X_0^\theta,\ldots,X_r^\theta$ are fuzzy sets.

By re-denoting input variables:

$(u_0(t), u_1(t), \ldots, u_m(t)) = (1, y(t-1), \ldots, y(t-r), x(t),\ldots, x(t-s))$,

finite difference equation's coefficients:

$$(c_0^\theta, c_1^\theta,\ldots,c_m^\theta) = (a_0^\theta, a_1^\theta,\ldots,a_r^\theta, b_1^\theta,\ldots,b_s^\theta),$$

and membership functions:

$$\left(U_1^\theta(u_1(t)),\ldots,U_m^\theta(u_m(t))\right) = \\ = \left(Y_1^\theta(y(t-1)),\ldots,Y_r^\theta(y(t-r)), X_0^\theta(x(t))\ldots,X_s^\theta(x(t-s))\right),$$

where $m=r+s+1$,

one obtains the analytic form of the fuzzy model (4), intended for calculating the output:

$$\hat{y}(t) = \mathrm{c}^T\tilde{\mathrm{u}}(t), \tag{17}$$

where:

$$\mathrm{c} = (c_0^1,\ldots,c_0^n,\ldots,c_m^1,\ldots,c_m^n)^T$$

is the vector of the adjustable parameters*;*

$$\tilde{u}^T(t) = (u_0(t)\beta^1(t),\ldots,u_0(t)\beta^\theta(t),\ldots \\ \ldots,u_m(t)\beta^1(t),\ldots,u_m(t)\beta^n(t))$$

is the extended input vector;

$$\beta^\theta(t) = \\ = \frac{U_1^\theta(u_1(t)) \otimes\ldots\otimes U_m^\theta(u_m(t))}{\sum_{\theta=1}^N\left(U_1^\theta(u_1(t)) \otimes\ldots\otimes U_m^\theta(u_m(t))\right)} \tag{18}$$

is a fuzzy function, where $\otimes$ denotes the minimization operation of fuzzy product.

If for $t = 0$, the vector $\boldsymbol{c}(0) = 0$, the correcting $nm\times nm$ matrix $Q(0)$ (m is the number input vectors, n is the number of production rules), and the values of $u(t)$, $t = 1,\ldots, N$ are specified, the parameter vector $\mathbf{c}(t)$ is calculated using the known multi-step LSM :

$$\mathrm{c}(t) = \mathrm{c}(t-1) + Q(t)\tilde{\mathrm{u}}(t)[y(t) - \mathrm{c}^T(t-1)\tilde{\mathrm{u}}(t)]$$

$$Q(t) = Q(t-1) - \frac{Q(t-1)\tilde{u}(t)\tilde{u}^T(t)Q(t-1)}{1+\tilde{u}^T(t)Q(t-1)\tilde{u}(t)}$$

$$Q(0) = \gamma I,\ \gamma >> 1,$$

where I is the unit matrix.

The above equations show that even in case of one-dimensional input and few production rules, a lot of observations are needed to apply LSM that makes the fuzzy model too unwieldy. Therefore, only a part of the whole set of rules ($r < n$) should be chosen according to a certain criterion.

The application of the associative search techniques, where one or more model parameters are fuzzy, is reduced to such determination of the predicate

$$\Xi = \{\Xi_i(R_0^a, R^a, T^a)\},$$

that the number of production rules in the TS model is significantly reduced according to some criterion. For example, the following matrix:

$$\begin{matrix} \beta_1^{\Theta_t} & \ldots & \beta_P^{\Theta_t} \\ \ldots & \ldots & \ldots \\ \beta_1^{\Theta_{t-s}} & \ldots & \beta_P^{\Theta_{t-s}} \end{matrix}$$

can be defined for P-dimensional input vectors at time steps t–j, $j = 1,\ldots,s$. If the rows of this matrix are ranged, say, w.r.t.

$$\sum_{p=1}^P\left|\beta_p^{\Theta_i}\right|$$

decrease and a certain number of rows are selected, then such selection combined with condition (4) will define the predicate

$$\Xi = \{\Xi_i(R_0^a, R^a, T^a)\}$$

and, respectively, the criterion for selecting the images (sets of input vector) from the history.

Let us range the rows of this matrix, for example, subject to the criterion of descending the values $\sum_{p=1}^P\left|\beta_p^{\Theta_i}\right|$, and select a certain number of rows. Such a selection combined with condition (4) defines the predicate $\Xi = \{\Xi_i(R_0^a, R^a, T^a)\}$, and, respectively, the image selection criterion (sets of input vectors) from the archive.

8 Fuzzy associative search

Notwithstanding all benefits delivered by fuzzy techniques, their application reduces significantly the calculations speed that is critical for predicting the dynamics of some plants. This consideration coupled with the principal impossibility of formalizing some factors necessitated the development of algorithms that could combine all advantages of fuzzy approach and associative search algorithms.

Assume the associative search procedure is determined by the predicate $\Xi(P^a,R^a)$, which interprets input variables' limits (specified, say, by process specifications) as a fuzzy conjunction of input variables:

$$\Xi(P^a,R^a)= \\ \{(X_1: x_1 \subset A_1) \wedge (X_2: x_2 \subset A_2)\ldots(X_n: x_n \subset A_n)\}$$

for all

$$X_1, X_2,\ldots X_n$$

from

$$DX = DX_1 \times DX_2 \times\ldots\times DX_n.$$

Then the production rules, where fuzzy variables possess such values that $\Xi(P^a, R^a)$ possesses the value FALSE, will be discarded automatically. This reduces drastically the number of production rules employed in the fuzzy model and thus increases significantly the algorithms' speed.

9 Fuzzy associative search based on clustering

The crucial problem of identification algorithms development on the basis of associative search is the selection at each time step of a new set of input vectors close to the current input w.r.t. some criterion. This is, in fact, process history data mining and, in particular, the cluster analysis task: automatic objects grouping, classification without teacher, or taxonomy. Data mining presumes information retrieval from historical process data and its representation in the form convenient for further analysis and control.

The problem of input vectors grouping in accordance with the criterion determined by associative impulse is solved. This is a multidimensional clustering problem.

Each object can be attributed to some group by assigning a cluster mark. A variety of cluster analysis algorithms are known [13].

In the associative problem, for selecting input vectors "close" to the current one, the cluster mark is determined by associative impulse, and the vectors are being sought inside the appropriate cluster. In the multi-dimensional case, the clustering task can be formulated as follows.

There is a sample of objects $l = \{\bar{x}_N, \dots \bar{x}_1\}$. For our purpose, we will consider input vectors of some process as such objects. At the learning phase, $K \geq 2$ clusters (groups of objects) will be formed. The number of clusters may be either determined in advance or be a solution to appropriate optimization problem.

We assume that each object $\bar{x}_i, i = 1,2, \dots N$ can be described by means of a set of variables ("characteristics")$x_{i1}, x_{i2}, \dots x_{iS}$. The set $\{x_{i1}, x_{i2}, \dots x_{iS}\}$ may contain the variables of various types, such as numerical, "qualitative" (or *categorical*), etc.

By *categorical data* we understand qualitative characteristics of objects measured in the name scale. When such scale is used, one has to denote only whether the objects are equal or not w.r.t. the measured attribute.

Let D_j denote the set of values of the variable x_{ij} for $\forall i = 1, \dots, N$. As$x_i = x_{i1}, x_{i2}, \dots x_{iS}$, we denote a set of observations for the object $\bar{x}_i$, $i = 1,2, \dots N$.

The set of variables' observations relevant to the sample will be represented as a data table V with N rows and S columns: $V = \{x_{ij}\}$, i=1,2,…,N, j=1,2,…,S; here the element x_{ij} located in the intersection of the i-th row and j-th column of the matrix corresponds to the observation of j-th variable for the i-th object. In our task, the formation of characteristics is determined by the process history; therefore the next phase, namely, metric selection for clusters formation, is the most important one.

10 Metric selection

The choice of metric is determined first of all by the objects space. There are a lot of metrics; the most popular ones are Euclidean, square Euclidean, Manhattan, exponential, Tchebyshev, and Mahalonobis.

1. *Euclidean distance* is the best choice in case of continuous real characteristics:

$$\rho(x_i, x_m) = \sqrt{\sum_{j=1}^{S}(x_{ij} - x_{mj})^2} = \|x_i - x_m\|_2,$$
$$i = 1,2, \dots N, m = 1,2, \dots N.$$

2. *Square of Euclidean Distance* is used for assigning larger weights to distant objects. It is defined as follows:

$$\rho(x_i, x_m) = \sum_{j=1}^{S}(x_{ij} - x_{mj})^2, \ i = 1,2, \dots N,$$
$$m = 1,2, \dots N.$$

3. *Manhattan Distance*. For this metric, the effect of separate large differences (outliers) is decreased because they are not squared:

$$\rho(x_i, x_m) = \sum_{j=1}^{S}|x_{ij} - x_{mj}|, i = 1,2, \dots N,$$
$$m = 1,2, \dots N.$$

4. *Tchebyshev Distance* is typically used when 2 objects are to be defined as "different", if they differ in some coordinate:.

$$\rho(x_i, x_m) = \max_{j=1}^{S}(|x_{ij} - x_{mj}|), i = 1,2, \dots N,$$
$$m = 1,2, \dots N.$$

5. *Exponential Distance* is applied when one needs to increase or decrease the weight, related with the coordinate in which the objects differ significantly:

$$\rho(x_i, x_m) = \sqrt[r]{\sum_{j=1}^{S}(x_{ij} - x_{mj})^p}, i = 1, \dots, N,$$
$$m = 1, \dots, N,$$

where r and p are determined by expert judgments.

6. Mahalonobis distance is used to exclude the effect of strong linear correlations between variables. It is defined as:

$$\rho^2(x_i, x_m) = (\bar{x}_i - \bar{x}_m)^T R^{-1}(\bar{x}_i - \bar{x}_m)$$

where R is the covariance matrix estimated per the sample or assumed to be known a priori.

Grouping may be crisp or fuzzy (grade of membership of each object to the groups is calculated).

11 Data Mining techniques in associative search tasks

The use of a priori information (process knowledge) plays the key role in the intelligent approach to predictive model building described above. Therefore, the intelligent data analysis, in particular, the selection of a data set meeting a certain criterion ("associative impulse") at each algorithmic step, plays the key role. In the simplest case, the associative impulse presumes the selection of the next vector from the history (this selection operation is called *association*) in such a way that this vector belongs to a certain domain in the space of vectors stored in the process history. A certain metric is introduced respectively.

Associative search algorithm discussed above is effective for the case, when the control plant is nonlinear, high response speed is not required, and the computational resources enable the exhaustive search of the process history. Such an approach is quite satisfactory for the identification of a rather broad class of control plants, for example, continuous and semi batch processes in chemical and oil refining industries.

However, the unpractical use of computational resources within this approach is obvious. Worse yet, each data search cycle requires to pick out the number of vectors sufficient for solving a system of equations, which, according to the LSM, allows to predict the output at the next time step. It is not guaranteed that such selection (even with high redundancy) is attainable in a single step.

12 Application of traditional clusterization methods for the associative search

To increase the algorithm speed (which is a key performance index of such algorithms for certain applications) and computational resource saving, it is proposed to teach the system. Clustering (learning without a teacher) looks an effective technique.

When using any of the known clustering algorithms ("crisp" case), the original set of objects $\overline{x}_i \in \mathrm{X}, i = 1, \ldots, N$ is split into several disjoint subsets. Here, any object from **X** belongs to a single class only.

When using fuzzy clustering techniques, it is allowed for one object to belong to several (or all) clusters simultaneously but with different degrees of certainty determined by the selected membership function. Here, the clusters are fuzzy sets. Fuzzy clustering may often be more preferable than crisp, for example, for the objects located at cluster borders.

C-Averages algorithm

Assume the number of clusters K is preassigned.

1. K points defined as "centers of gravity" are selected in a random way.
2. Each object is referred to a cluster with the nearest "center of gravity"
3. "Centers of gravity" are recalculated subject to the previous operation.

In the case of crisp clustering, the c-averages algorithm splits the set **X** into subsets $\mathbf{A}_k$, $k = 1, \ldots, K$, and the following requirements are met:

$\bigcup_{k=1}^{K} \mathrm{A}_k = \mathrm{X}$ – each object should be attributed to a certain cluster;

$\mathrm{A}_k \cap \mathrm{A}_r, k, r = 1, \ldots, K$ – each object belongs to one and only one cluster;

$\varnothing \subset \mathrm{A}_k \subset \mathrm{X}, k = 1, \ldots, K$ – no cluster can be empty or contain all objects.

In process of clustering, a characteristic function is used which may take on 2 values: 0 if an element does not belong to a cluster, and 1 if it does. The clusters can be described by the following decomposition matrix:

$$\mathrm{U} = [u_{ki}], k = 1, \ldots, k; i = 1, \ldots, N,$$

where the k-th row of the matrix **U** denotes that the objects $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_N$ belong to the cluster $\mathrm{A}_k, k = 1, \ldots, K$.

$$u_{ki} \in \{0,1\}, k = 1, \ldots, k; i = 1, \ldots, N.$$

The matrix **U** should have the following properties:

$$\sum_{k=1}^{K} u_{ki} = 1, i = 1, \ldots, N.$$

$$0 < \sum_{i=1}^{N} u_{ki} \leq N - 1, k = 1, \ldots, K.$$

To estimate decomposition quality, a scatter criterion is used which represents the sum of distances between the objects and the center of their cluster. For Euclidean space, this criterion looks as follows:

$$\sum_{k=1}^{K} \sum_{x_i \in \mathrm{A}_k} \|g_k - x_i\|^2,$$

where $\bar{x}_i \in \mathrm{X}, i = 1, \ldots, N$ is the *i*-th object of clustering, $\mathbf{A}_k = \|\bar{x}_p\|, u_{kp} = 1, p = 1, \ldots, N$ is the cluster with the number *k*,

$$g_k = \frac{1}{|\mathrm{A_k}|}\sum_{x_i \in \mathrm{A}_k}\|x_i\|$$

is the center of the *k*-th cluster.

Fuzzy c-averages algorithm

This algorithm comprises the following sequence of operations.

Initial decomposition of *N* objects into *K* clusters is determined by selecting the membership matrix $\mathrm{F} = [\mu_{ki}], k = 1, \ldots, k; i = 1, \ldots, N$. Typically, $\mu_{ki} \in [0,1]$ are selected, where the *k*-th row of the matrix **F** denotes that the objects $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N$belong to the cluster $\mathrm{A}_k, k = 1, \ldots, K$.

The only difference between the matrices **F** and **U** is that in the fuzzy decomposition, the grade of object's membership in a cluster may take any value between 0 and 1, while in the crisp one, it may equal either 0 or 1.

Similar conditions for fuzzy decomposition matrix are defined:

$$\sum_{k=1}^{K} \mu_{ki} = 1, i = 1, \ldots, N,$$
$$0 < \sum_{i=1}^{N} \mu_{ki} \le N - 1, k = 1, \ldots, K.$$

Fuzzy decomposition enables, in particular, a simple solution to the problem of objects located at the border between 2 clusters: grades of membership equal to 0.5 are assigned to both. The following scatter criterion is used for estimating the decomposition quality [14]:

$$\sum_{k=1}^{K}\sum_{i=1}^{S}\|g_k - x_i\|^2 \mu_{ki}^m \quad ,$$

where $\bar{x}_i \in \mathrm{X}, i = 1, \ldots, N$ is the *i*-th object of clustering,

$$\mathbf{A}_k = \|\bar{x}_p\|, u_{kp} = 1, p = 1, \ldots, N$$

is the cluster with the number *k*,

$$g_k = \frac{\sum_{i=1}^{S} \mu_{ki}^m \|x_i\|}{\sum_{i=1}^{S} \mu_{ki}^m}$$

is the center of the *k*-th fuzzy cluster, $m \in [1, \infty)$ is the exponential weight factor describing clusters' fuzziness.

13 Solving the associative search problem by means of clusterization techniques

Associative search problem is solved by clustering technique (both crisp and fuzzy) in the following way.

The current vector under investigation is attributed to a certain cluster per the criterion of minimum distance to the center:

$$\min_k \sum_{k=1}^{K} \|g_k - \bar{x}_N\|^2,$$

where $\bar{x}_N \in \mathrm{X}$ is the current input vector of the control plant under investigation.

Further, the vectors are picked out from this cluster, which meet the selected associative search criterion. If the vectors picked out from the cluster are not enough for solving the prediction problem by means of LMS, the cluster can be enlarged by one of the known single-link methods which combine 2 clusters with minimum distance between any 2 of their members.

This naturally ensures a huge saving of computational resources as against the exhaustive search over the whole process history at each step. However, this aggregation into a new cluster presumes that many objects deliberately not meet the associative search criterion.

The approach described below looks the most reasonable.

Virtual clustering ("impostor" method)

For each time step *N*, the current vector under investigation is attributed to a certain cluster per the criterion of the minimum distance to the center – in the same way as with the traditional method.

Assume

$$\min_k \sum_{k=1}^{K} \|g_k - \bar{x}_N\|^2$$

is attained for *k*=*r*. Further, $\bar{x}_N$ is assigned to be the center of the cluster $\mathbf{A}_r$. If additional selection is needed from the archive of vectors meeting the associative search criterion, then the clusters are chosen for aggregation with the minimum distance between their centers and the vector $\bar{x}_N$. In such case not only a considerable number of vectors distant from $\bar{x}_N$ will be discarded, but also the maximum possible number of vectors meeting the associative search criterion will appear.

After the associative search is completed, the assignment of $\bar{x}_N$ as the center of the cluster $\mathbf{A}_r$ is cancelled, and the process is continued in the same way as in the traditional algorithm.

14 The methodology of the intelligent VSCS synthesis

The sets of values of signals, forming virtual model (9) in different time instants, and values of $s(t)$, corresponding to them, fill in the Knowledge Base of the system (see Fig. 1). Under accounting the content of the knowledge base, involving values of the signal $\sigma(t)$ one may obtain (by use of the LS method) statistically effective unbiased prediction of the system output. However, under the closed-loop description, it is reasonable the operator **D**, whose main function is to generate the system dynamics, to implement the multiplication of $e(t)$ and values of a limited signal $d(t)$, where $e(t)$ and $d(t)$ are stochastically independent.

The virtual model that is synthesized by virtue of changing the internal system dynamics is the generalized information model of the control plant.

In the event, when data on the coordinate signals are taken into account only, that is in the model coefficients at e, σ are set to be equal to zero, we come to the conventional closed-loop control scheme with an identifier accompanied with all its problems.

In the event, when, wise versa, in the model the coefficients at e, σ are not equal to zero, while the coefficients at g, y, u are equal to zero, we obtain the information model of the internal system dynamics. Such a model will provide vanishing the dynamic error e, σ. However, in the general case, meeting certain control properties is not guaranteed, and, first of all, the stability. In other words, the information plant model permits not only to provide vanishing the dynamic error, but also to keep certain quality of the system performance.

In papers [8, 9], sufficient criteria of the stability of time-varying dynamic plants in the terms of the spectrum of the multi-scale wavelet expansion of inputs and outputs of the system. And since under deriving virtual predicting models the associative search procedure is used, one may say about the method of the synthesis of VSCS for a broad class of non-linear systems.

More flexible control (in particular, providing the stability), achieved by use of the system dynamics generator, may be obtained due to using the generalized information model under forming $s(t)$ and **S**.

Forming the information model of kind (3), in particular, we obtain the associative virtual model of the dynamics generator.

15 Conditions of the associative model stability in the aspect of the analysis of the spectrum of multi-scale wavelet expansion

Let a predicting associative model of a non-linear time-varying plant meet equation (3).

For the selected detail level L for the current input vector $x(t)$, we obtain the multi-scale expansion [9]:

$$\begin{aligned} x(t) &= \sum_{k=1}^{N} c_{L,k}^{x}(t)\varphi_{L,k}(t) + \\ &+ \sum_{l=1}^{L}\sum_{k=1}^{N} d_{l,k}^{x}(t)\psi_{l,k}(t), \\ y(t) &= \sum_{k=1}^{N} c_{L,k}^{y}(t)\varphi_{L,k}(t) + \\ &+ \sum_{l=1}^{L}\sum_{k=1}^{N} d_{l,k}^{y}(t)\psi_{l,k}(t), \end{aligned} \tag{19}$$

where: L is the depth of the multi-scale expansion ($1 \le L \le L_{max}$, where $L_{max} = \lfloor \log_2 N^* \rfloor$ and N^* is the power of the set of states of the system in the System Dynamics Knowledge Base); $\varphi_{L,k}(t)$ – are scaling functions; $\psi_{l,k}(t)$ are the wavelet functions that are obtained from the mother wavelets by the tension/combustion and shift

$$\psi_{l,k}(t) = 2^{l/2}\psi_{\text{mother}}\left(2^{l}t - k\right)$$

(as the mother wavelets, in the present case we consider the Haar wavelets); l is the level of data detailing; $c_{L,k}$ are the scaling coefficients, $d_{l,k}$ are the detailing coefficients. The coefficients are calculated by use of the Mallat algorithm [10].

Let us expand equation (3) over wavelets:

$$\begin{aligned} &\sum_{k=1}^{N} c_{Lk}^{y}(t)\varphi_{Lk}(t) + \sum_{l=1}^{L}\sum_{k=1}^{N} d_{lk}^{y}(t)\psi_{lk}(t) = \\ &= \sum_{k=1}^{N}\left(\sum_{i=1}^{m} a_i c_{Lk}^{y}(t-i)\varphi_{Lk}(t-i)\right) + \\ &+ \sum_{l=1}^{L}\sum_{k=1}^{N}\left(\sum_{i=1}^{m} a_i d_{lk}^{y}(t-i)\psi_{lk}(t-i)\right) + \\ &+ \sum_{k=1}^{N}\left(\sum_{s=1}^{S}\sum_{j=1}^{r_s} b_{sj}\, c_{Lk}^{s}(t-j)\varphi_{Lk}(t-j)\right) + \\ &+ \sum_{l=1}^{L}\sum_{k=1}^{N}\left(\sum_{s=1}^{S}\sum_{j=1}^{r_s} b_{sj}\, d_{lk}^{s}(t-j)\psi_{lk}(t-j)\right). \end{aligned}$$

Let us consider individually the detailing and approximating parts (12) and (13) correspondingly:

$$d_{lk}^{y}(t)\psi_{lk}(t) = \tag{20}$$

$$= \sum_{i=1}^{m} a_i d_{lk}^{y}(t-i)\psi_{lk}(t-i) + \\ + \sum_{s=1}^{S}\sum_{j=1}^{r_s} b_{sj}\, d_{lk}^{s}(t-j)\psi_{lk}(t-j), \\ c_{Lk}^{y}(t)\varphi_{Lk}(t) = \\ = \sum_{i=1}^{m} \hat{a}_i c_{Lk}^{y}(t-i)\varphi_{Lk}(t-i) + \\ + \sum_{s=1}^{S}\sum_{j=1}^{r_s} \hat{b}_{sj}\, c_{Lk}^{s}(t-j)\varphi_{Lk}(t-j). \quad (21)$$

In papers [8, 9] it was shown that a sufficient condition of the stability of plant (3) (and, hence, also (11) is as follows: for $\forall k = \overline{1, N}$ meeting the inequalities is to be provided:

– if $m > R$, $R = \max_{s=\overline{1,S}} r_s$, then the condition for the detailing coefficients:

$$\left|a_m d_{lk}^{y}(t-m)\right| < \left|d_{lk}^{y}(t)\right|,$$

for the approximating coefficients:

$$\left|a_m c_{Lk}^{y}(t-m)\right| < \left|c_{Lk}^{y}(t)\right|,$$

– if $m < R$, $R = \max_{s=\overline{1,S}} r_s$, then the condition for the detailing coefficients:

$$\left|\sum_{s=1}^{S} b_{sR} d_{lk}^{s}(t-R)\right| < \left|d_{lk}^{y}(t)\right|,$$

for the approximating coefficients:

$$\left|\sum_{s=1}^{S} b_{sR} c_{Lk}^{s}(t-R)\right| < \left|c_{Lk}^{y}(t)\right|,$$

– if $m = R \neq 1$, $R = \max_{s=\overline{1,S}} r_s$, then the condition of the stability for the detailing coefficients:

$$\left|a_m d_{lk}^{y}(t-m) + \sum_{s=1}^{S} b_{sm} d_{lk}^{s}(t-m)\right| < \left|d_{lk}^{y}(t)\right|,$$

for the approximating coefficients:

$$\left|a_m c_{Lk}^{y}(t-m) + \sum_{s=1}^{S} b_{sm} c_{Lk}^{s}(t-m)\right| < \left|c_{Lk}^{y}(t)\right|,$$

– if $m = R = 1$, $R = \max_{s=\overline{1,S}} r_s$, then the condition of the stability for the detailing coefficients:

$$\left|a_1 d_{lk}^{y}(t-1) + \sum_{s=1}^{S} b_{s1} d_{lk}^{s}(t-1)\right| < \left|d_{lk}^{y}(t)\right|,$$

for the approximating coefficients:

$$\left|a_1 c_{Lk}^{y}(t-1) + \sum_{s=1}^{S} b_{s1} c_{Lk}^{s}(t-1)\right| < \left|c_{Lk}^{y}(t)\right|.$$

The investigation of the stability by use of the wavelet analysis will be demonstrated by use of the example of plant (10) described in Section 6. The model has the input and output memory depth $m = n = 1$. For this model, criteria of the stability for the approximating and detailing coefficients will have the form:

$$\left|\frac{a_1 c_{Lk}^{S}(t-1) + b_1 c_{Lk}^{Ia}(t-1) + b_2 c_{Lk}^{Ua}(t-1)}{c_{Lk}^{S}(t)}\right| < \\ < 1,$$

$$\left|\frac{a_1 d_{jk}^{S}(t-1) + b_1 d_{jk}^{Ia}(t-1) + b_2 d_{jk}^{Ua}(t-1)}{d_{jk}^{S}(t)}\right| < \\ < 1,$$

$$1 \le j \le L \le L_{max}, \; k = \overline{1, N}.$$

In Fig. 3-6, meeting the stability criterion is displayed for the approximating and detailing coefficients of the sample for the prediction based on the associative search, where L is the depth of the expansion ($L_{max} = 14$), and N is the quantity of vectors selected to derive the approximating model by use of the associative search. The simulation results shows that for a number of vectors selected the stability criterion is not met for the prediction. Thus, in these instants a non-stationarity is present that requires additional studying.

Investigating expansion coefficients (11) in accordance to the wavelet stability criterion, in entity, provides not only solving the inverse problem of the spectral analysis for the non-linear operator $\boldsymbol{D}$, but also permits to select such control algorithm that it will keep the system stability. Just in such a manner one should form the operator $\boldsymbol{D}$ under solving the problem of the synthesis of the variable structure systems.

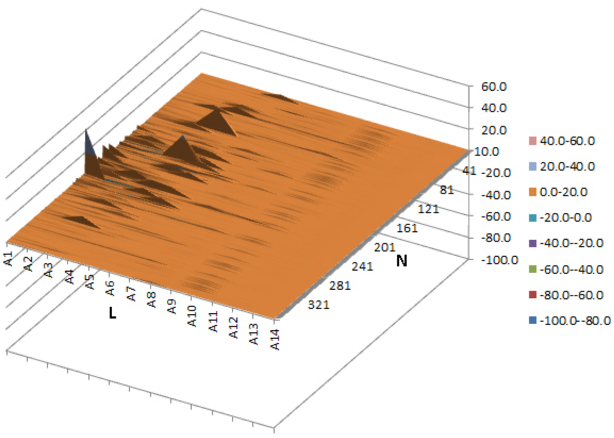

Fig. 3. Stability criterion for the approximating coefficients (general view)

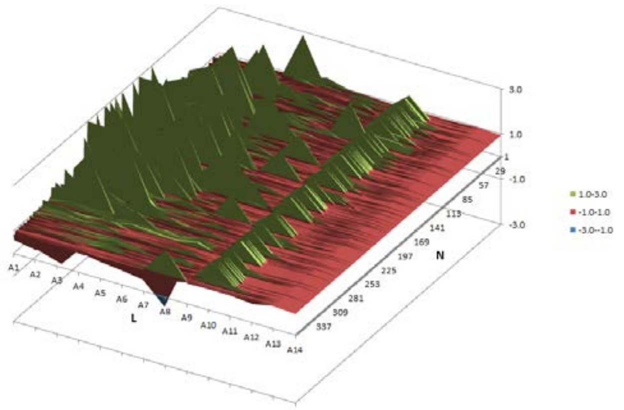

Fig. 4. Stability criterion for the approximating coefficients (enlarged view)

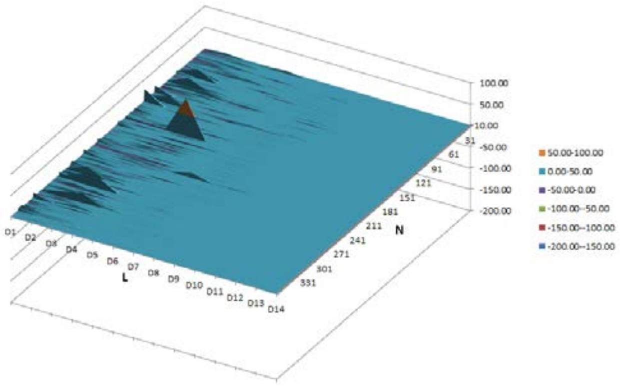

Fig. 5. Stability criterion for the detailing coefficients (general view)

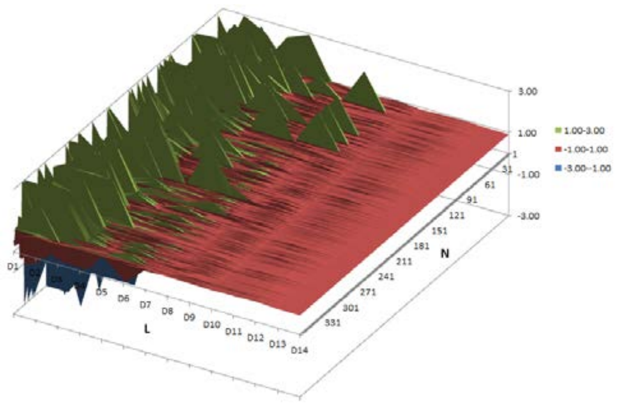

Fig. 6. Stability criterion for the detailing coefficients (enlarged view)

16 Conclusions

For non-linear and time-varying dynamic systems, a methodology of the variable structure systems sybthesis is proposed, based on deriving operator feed-backs and virtual models of the plant dynamics. The predicting virtual models are based on the intelligent data analysis of the dynamic dossier of the system status. The application of this approach is especially effective whilst compensating for insufficient lab data for model development. In such a case, fuzzy specification of certain process variables using process knowledgebase is practiced.

On the basis of approach proposed, it looks possible to synthesize systems meeting set dynamic properties. It is predicted approaching to the stability bounds on the basis of investigating the dynamics of the coefficients of the multi-scale wavelet analysis.

References:

[1] S. Emelyanov, Variable Structure Systems as a Gateway to New Types of Feedback Structures, *IFAC Proceedings Volumes (IFAC-PapersOnline). World Congress*, Vol. 18, No. 1, 2011, pp. 747-755.

[2] N.N. Bakhtadze, V.V. Kulba, I.B. Yadikin, V.A. Lototsky, E.M. Maximov, Identification methods based on associative search procedure, *Management and Production Engineering Review*, Vol. 2, No. 3, 2011, pp. 6-18.

[3] N.N. Bakhtadze, V.A. Lototsky, E.M. Maximov, Associative Search method in system identification, *Proceedings of the 14th International conference on Automatic control, Modeling and Simulation*, Saint Malo & Mont Saint Michel, France, 2012, pp. 49-57.

[4] N. Bakhtadze, A. Dolgui, V. Lototsky, E. Maximov, Intelligent Identification Algorithms for Frequency/Power Control in Smart Grid, *IFAC Proceedings Volumes (IFAC-PapersOnline). Information Control Problems in Manufacturing*, Vol. 14, No. 1, 2012, pp. 940-945.

[5] A.L. Bunich, Degenerate linear-quadratic problem of discrete plant control under uncertainty, *Automation and Remote Control*, Vol. 72, No. 1, 2011, pp. 2351-2363.

[6] E.H. Moore, On the reciprocal of the general algebraic matrix, *Bulletin of the American Mathematical Society*, Vol. 26, 1920, pp. 394-395.

[7] R. Penrose, A generalized inverse for matrices, *Proceedings of the Cambridge Philosophical Society*, Vol. 51, No. 3, 1955, pp. 406-413.

[8] N. Bakhtadze, V. Lototsky, S. Vlasov, E. Sakrutina, Associative Search and Wavelet Analysis Techniques in System Identification, *IFAC Proceedings Volumes (IFAC-PapersOnline). System Identification*, Vol. 16, No. 1, 2012, pp. 1227-1232.

[9] N.N. Bakhtadze, E.A. Sakrutina, The Intelligent Identification Technique with Associative Search, *International Journal of*

Mathematical Models and Methods in Applied Sciences, Vol. 9, 2015, pp. 418-431.

[10] S. Mallat, *A Wavelet Tour of Signal Processing.* (Third edition) The Sparse Way, Elsevier, Amsterdam, 2009, 805 p.

[11] S. Wu, M.J. Er, Y. Gao, A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks, *Fuzzy Systems, IEEE Transactions*, Vol. 9, No. 4, 2003, pp. 578-594.

[12] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Trans. Systems. Man Cybernetics*, Vol. 15, No. 1, 1985, pp. 116-132.

[13] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review. *ACM Computing Surveys*, Vol. 31, No. 3, 1999, pp. 264-323.

[14] R.D. Pascual-Marqui, A.D. Pascual-Montano, K. Kochi, J.M. Carazo, Smoothly distributed fuzzy C-means: a new self-organizing map, *Pattern Recognition*, Vol. 34, 2001, pp. 2395-2402.