Management and Production Engineering Review
Volume 2 • Number 3 • September 2011 • pp. 6–18

IDENTIFICATION METHODS BASED ON ASSOCIATIVE SEARCH PROCEDURE

Nataliya N. Bakhtadze*, Vladimir V. Kulba, Igor B. Yadikin, Vladimir A. Lototsky, Evgeny M. Maximov

V.A. Trapeznikov Institute of Control Sciences, Moscow, Russia

Corresponding author:
Natalia N. Bakhtadze
V.A. Trapeznikov Institute of Control Sciences
65 Profsoyuznaya St., 17997 Moscow, Russia
phone: +74 95 334-92-01
e-mail: sung7@yandex.ru

Received: 20 April 2011
Accepted: 19 August 2011

Abstract

In modern control systems, ***identification*** is an integral part of adaptive control where process models are adjusted using real-time operation data and control actions optimal with respect to some performance criterion are developed. A variety of identification methods based on mathematical statistics techniques have been developed. Algorithms optimal for certain classes of objects and external disturbances were categorized dependent on the available a priori information about the control object. The limits of approximating models development and application were outlined.

Against this background, the paper presents novel associative search techniques enabling the development of a new dynamic object's model on each time step rather than plant approximation pertaining to time. The model is build using the data samples from process history (associations) developed at the learning phase. The new techniques employs the models of human individual's (process operator's, stock analyst's or trader's) behavior based on professional knowledge formalization. Application examples from oil refining and chemical industries, power engineering, and banking are adduced.

Keywords

process identification, knowledgebase, associative search models, soft sensors.

Introduction

Strict theoretical results in adaptive control of dynamic plants were obtained in the beginning of the 1980s [1, 2]. The identifiers typically used least mean squares technique and its modifications (Goodwin's scheme [3]). Later, the progress was achieved in optimizing the control strategies with respect to (w.r.t.) control cost function for the plants both with parametric and non-parametric uncertainty [4].

The theoretic results were exemplified in appropriate software products compatible with MATLAB (System Identification Toolbox); they were based on popular algorithms [5].

A*dvanced process control systems* are extensively used in the growing number of industrial applications. The concept of advanced process control (APC) comprises a variety of techniques and approaches from advanced regulatory controls (also known as "traditional advanced controls" such as feedforward, ratio control, pass balancing, Smith predictors, etc.) through multivariable *model predictive control* (MPC) and real-time optimization (RTO) of several process units. Owing to predictive models, MPC applications make it possible to decrease the variability of process variables and, hence, to operate process units tighter against constraints that results in higher throughput, yields improvement, energy conservation, lower operating expenses, and other benefits. An MPC system typically includes a set of soft sensors that allows to control product qualities directly as well as consider them as constraints in process optimization. Modern MPC systems feature sophisticated soft sensor development techniques based on applied statistics, adaptive approaches and "first-principles" (hi-fi) modeling.

In modern decision-making support systems, the identification is done aiming to investigate the process properties that result in optimal control decision-making by human operator. *The prediction model technique* is one of so-called substitution methods: the unknown parameter is implicitly determined as the unique point minimizing a cost functional which is substituted by empirical one through the estimates calculation.

The use of stochastic approximation algorithms (both gradient and pseudo-gradient) for minimizing the empirical functional make it possible to adjust the model in real time. For example, with quadratic cost function we get generalized recursive LMS technique, while slower increasing cost functions result in different estimation algorithms, robust w.r.t. innovation distribution [6].

The predictive model technique can be effectively implemented for Gaussian ARMA disturbances. For non-Gaussian disturbances, the optimal available control strategy is generally formed by nonlinear feedback [7], being evidently a much harder exercise.

Against this background, the paper presents the identification technique based on *virtual models* design. The term "virtual" should be understood as *"ad hoc"*. This technique named *"associative search"* suggests predicting model design for dynamic plant on each time step using the sets of historical data (associations) obtained at the learning stage rather than real process approximation in time. Such approach is close to idea of employing additional a priori information about the control plant in learning process [8].

Fuzzy logic technique is used to develop the algorithms. Thus, the proposed methods imply for real-life plant identification to use the simulation of human operator behavior based on process knowledge formalization.

Modern soft sensors in industrial systems

Software systems named **soft sensors** (or inferential calculations) implement the approach to model building based on identification analysis. When building a model of a specific process, they employ the models from other control levels along with the current and historic data. It is important here that the models of various production cells do not become the elements of a more complicated model at a higher level but rather supply the values to the input information vector of that model. In fact, the lack of a priori information about the process under investigation is compensated for by what may be called additional, or "virtual", measurements.

Modern soft sensors employ a variety of traditional data analysis and control theory techniques as well as neuron networks, fuzzy logic, and genetic algorithms.

Soft sensors are typically used for supporting process operators' decision-making because they enable on-line product quality prediction. By "operator" we mean either process operators from various industries or stock analysts and traders. Recommended control actions are presented to operators directly from process schematics or through an independent interface. Identification algorithms underlying modern soft sensors are based on expert knowledge. Soft sensors employ both decision-maker's expert knowledge and process knowledge bases. In the second case, the operator is offered either the recommended control action or the values of process variables obtained by means of process monitoring.

Soft sensors based on process technology knowledge implement the **intelligent approach to identification models design**.

A method using prediction models based on the imitation of analyst's associative thinking can be considered. Identification algorithms employed in modern control systems often use expert knowledge both from human expert and from a knowledgebase. In the second case, an operator can choose between recommended control action and a forecast based on process state monitoring.

Two knowledge types are distinguished: declarative and procedural [9]. The first type includes the description of various facts, events, and observations, while skills and experience refer to the second type. Experts differ from novices by the structure and way of their thinking and, in particular, the searching strategy [10]. If a person is not experienced, he/she would use the so-called "backward reasoning": review different possible answers and makes a decision in favor of a specific answer based on the information received from the process at the current time step. On the contrary, an expert does not need to analyze current information in the process of decision-making, rather he/she uses the so-called 'forward reasoning' method, which implies that the decision-making strategy is created subconsciously and this strategy is nonverbal. Therefore, in terms of the method of computational view of thought [11] system effectiveness will to a great extent be determined by expert's qualification and by the available a priori information. Within the framework of this method, the cognitive psychology determines knowledge as a certain set of actually existing elements-symbols stored in human memory, processed during thinking and determining the behavior. The symbols, in turn, could be determined by their structure and the nature of neuron links [12].

Knowledge processing in an intelligent system consists in the recovery (associative search) of knowledge by its fragment [14]. The knowledge can be defined as an associative link between images. The associative search process can take place either as a process of image recovery using partially specified symptom (or knowledge fragment recovery by incomplete information; this process is usually emulated in various associative memory models) or as searching others images (linked associatively with the input image) related with other time steps. Those images make sense as a cause or an effect of an input image.

In [14] the model is offered which describes the associative thinking process as a sequential process of remembering based on *associations* – pairs of images defined by a set of symptoms. Such model can be considered as an intermediate level between neuron network models and logical models used in classical artificial intelligence systems. Further, we discuss an approach to developing on-line support of human individual's (process operator's or trader's) decision-making based on the dynamic simulation of associative search and the identification technique based of virtual models.

Various associative search schemes are known [13]. For example, in frame systems, the search problem is implemented as frame matching. In semantic networks, the search is performed by comparing network fragments and the query graph. In discrete multi-objective choice tasks, an approach based on *verbal decisions analysis* techniques [15] has demonstrated its effectiveness. It is based on objects' description decomposition subject to several criteria into partial descriptions of smaller dimensions that are further offered to the decision-maker for comparison (with the assumption of pairwise equal estimates for the criteria not included in the descriptions).

In [14] a model is offered that describes the associative thinking as a sequential process of recalling by means of *associations* – the pairs of images each characterized by a set of symptoms. This model is an intermediate one between the neural network and the logical models used in classical artificial intelligence systems.

The paper offers an approach to the development of decision-making support for process operators by building an associative search procedure model at every time step.

An identification algorithm for complex nonlinear dynamic objects such as continuous and batch processes was presented in [16]. The identification algorithm with continuous real-time self-tuning is based on virtual models design.

At every time step, a new virtual model is created. To build a model for a specific time step, a temporary "ad hoc" database of historic and current process data is generated. After calculating the output forecast based on object's current state, the database is deleted without saving.

The linear dynamical prediction model looks as follows:

$$y_t = a_0 + \sum_{i=1}^{r} a_i y_{t-i} + \sum_{j=1}^{s} \sum_{p=1}^{P} b_{jk} x_{t-j,p}, \qquad (1)$$

where y_t is the object's output forecast at the t-th step, x_t is the input vector, r is the output memory depth, s is the input memory depth, P is the input vector length.

The original dynamic algorithm consists in the design of an approximating hypersurface of input vector space and the related one-dimensional outputs at every time step (see Fig. 1). To build a virtual model for a specific time step, the points close in a manner to the current input vector are selected. The output value at the next step is further calculated using least mean squares (LMS).

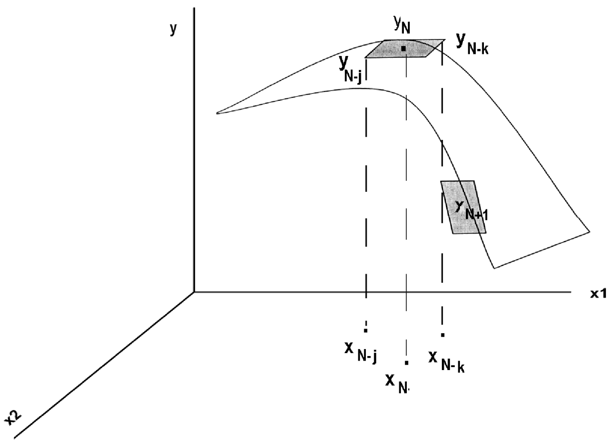

Fig. 1. Approximating hypersurface design.

Associative search technique for virtual models design

Consider a way of approximating hypersurface design. We use a method based on the associative thinking model. High-speed approximating hypersurface design algorithms enabling the usage of fuzzy models for various process applications were offered [17].

The following quantities

$$d_{t,t-j} = \sum_{p=1}^{P} |x_{tp} - x_{t-j,p}|, \qquad j = 1, ..., s, \quad (2)$$

were introduced as distances (metric in $\Re^P$) between points of P-dimensional input space, where, generally, $s < t$, and x_{tp} are the components of the input vector at the current time step t.

Assume that for the current input vector $\mathbf{x}_t$:

$$\sum_{p=1}^{P} |x_{tp}| = d_t. \quad (3)$$

To build an approximating hypersurface for $\mathbf{x}_t$, we select such vectors $\mathbf{x}_{t-j}$, $j = 1, \ldots, s$ from the input data archive such that for a given D_t the following condition will hold:

$$d_{t,t-j} \le d_t + \sum_{p=1}^{P} |x_{t-j,p}| \le d_t + D_t, \quad j = 1, ..., s. \quad (4)$$

The 2-D case is illustrated below (Fig. 2).

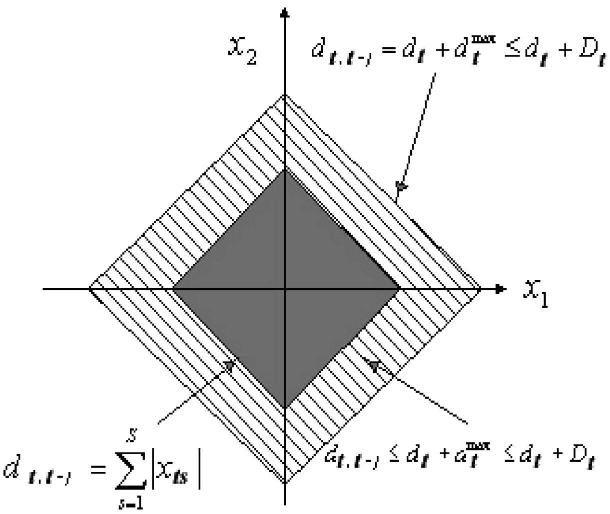

Fig. 2. An example of 2-D approximating hypersurface.

The preliminary value of D_j is determined on the basis of process knowledge. If the selected domain does not contain enough inputs for applying LMS, i.e., the corresponding SLAE has no solution, then the chosen points selection criterion can be slackened by increasing the threshold D_t.

To increase the speed of the virtual models-based algorithm, an approach is applied based on employing a model of process operator's or trader's associative thinking for predicting.

For modeling the associative search procedure imitating the intuitive prediction of process status by an analyst (we assume that the sets of process variable values, which are the components of an input vector, as well as the system outputs at previous time steps altogether create a set of symptoms, making an image of the object output at the next step.

The associative search process consists in the recovery of all symptoms describing the specific object based on its images. Denote the image initiating the associative search by R_0 and the corresponding resulting image of the associative search by R. A pair of images (R_0, R) will be further called association A or $A(R_0, R)$. The set of all associations over the set of images forms the memory of the knowledgebase of the intelligent system.

At the system's learning phase, an archive of images is created. In our case, a set of n input vectors selected form the process history according to the algorithm described in Sec. 1 will be considered as an image. At the prediction stage, the input vector x_t will be considered as an initial image R_0^a of the associative search, while approximating hypersurface formed by the input vectors from the process history built with the help of the algorithm from Sec. 1 will be the final image R^a of the associative search. This means that to build a virtual model, one should select the existing hypersurfaces stored in the archive at the learning phase rather than individual vectors close to x_t. The selected hypersurface is an image of the current input vector which is used for output prediction. The algorithm implements the process of image R^a recovery based on R_0^a, i.e., the associative search process, and can be described by a predicate $\Xi = \{\Xi_i(R_{0i}^a,\ R_i^a,\ T^a)\}$ where $R_{0i}^a \subseteq R_0$, $R_i^a \subseteq R$, and T^a is the duration of the associative search.

For the algorithm above, this predicate is a function asserting the truth or the falsity of input vector's membership of the specific domain in the inputs space. Therefore, the associative search process is reduced to the selection of a certain set of input vectors satisfying the condition (4) from the process archive. If the process history contains no image satisfying (4), then either the threshold D_t should be increased, or for a certain image of our input vector, some symptom should be replaced with a more relevant one. Formally, this means that the "worst" (i.e., located farther away from the current input then the rest ones w.r.t. the chosen criterion) vector from the process history will be deleted and replaced with a more relevant one, and so on.

Therefore, the analyst's decision-making process at any time step t could be constructed as associative search (process of remembering) of images (similar situations). The coordinates of approximating hypersurfaces used at previous steps are kept in archive.

Associative impulses

The criterion of input data selection from process history is called *associative impulse*. The following four basic groups of associative impulses can be defined dependent on the specification of the predicate, which defines the criterion.

1. The case where input vectors belong to some domain in the space $\Re^n$ looks the most natural for practical tasks. In this case, the associative impulse will be called *interval associative impulse*.
2. If in the expression of the predicate Ξ the attribute values are defined with the help of certain functional relationships (such as, e.g., $\Xi = \{x_1^2 = 4; x_1 + 2x_2 = 6\}$), then the associative impulse is called functional associative impulse.
3. If the predicate's arguments are the results of some logical operations, we call this a logical type associative impulse.
4. Finally, if predicate's argument is a fuzzy variable, the associative impulse will be referred to as fuzzy.

Associative search procedures in short-term forecasting

More sophisticated models can be proposed for the processes featuring complex nonlinear dynamics. Their design is further discussed with the case study of short-term stock prices at a stock exchange. In short-term price prediction (for trading process), not only the current situation but also the price dynamics is very important. The conventional regression models are not precise enough to handle this problem.

We apply the associative search procedure with more complicated requirements to approximating hypersurface selection [18]. We select from the archive such hypersurface (corresponding to some $\mathbf{x}_{t-j}$, $j = 1, ..., s$), that (i) it contains input vector at the current time step t, and (ii) the hypersurface corresponding to $\mathbf{x}_{N-j-1}$, $j = 1, ..., s$ contains the input vector at the previous time step $t-1$. Formally, this means that the predicate becomes more complex:

$$\Xi(R_0^a, R^a, T^a) = \left\{ \sum_{p=1}^{P} |x_{t-j,p}| \le D_t - \sum_{p=1}^{P} |x_{tp}|; \right. \\ \left. \sum_{p=1}^{P} |x_{t-j-1,p}| \le D_{t-1} - \sum_{p=1}^{P} |x_{tp}| \right\}. \tag{5}$$

There is principal opportunity to find more precise rules in the process of price changing by increasing the memory, say, to l steps ($l < t$).

$$\Xi(R_0^a, R^a, T^a) = \\ \left\{ \sum_{p=1}^{P} |x_{t-j,p}| \le D_t - \sum_{p=1}^{P} |x_{tp}| \sum_{p=1}^{P} |x_{t-j-1,p}| \le D_{t-1} - \right. \\ \left. \sum_{p=1}^{P} |x_{tp}|\,; ..., \sum_{p=1}^{P} |x_{t-l,p}| \le D_{t-l} - \sum_{p=1}^{P} |x_{tp}| \right\}. \tag{6}$$

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 index's 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 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.

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

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

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 rulebase in the fuzzy Mamdani system is a set of fuzzy rules such as:

$$R_j : LX_{1,j_1} \text{ AND } \ldots \text{ AND } LX_{n,j_n} \to LY_j. \tag{7}$$

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

$$R_j : LX_{1,j_1} \text{ AND } \ldots \text{ AND } LX_{n,j_n} \rightarrow r_j, \tag{8}$$

where r_j is real number to estimate the output y.

The j-th rule in Takagi-Sugeno model [20] looks as follows:

$$\begin{aligned} &R_j : LX_{1,j_1} \text{ AND } \ldots \text{ AND } LX_{n+m,j_{n+m}} \rightarrow \\ &r_{0j} + r_{1j}x_1 + r_{2j}x_2 + \ldots + r_{(n+m)j}x_{n+m}, \end{aligned} \tag{9}$$

where the output y is estimated by a linear function.

Thus, the fuzzy system performs the mapping $L : \Re^{n+m} \rightarrow \Re$.

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

$$\begin{gathered} \frac{\sum\limits_{i=1}^{\mathcal{K}} |f(\mathbf{x}_i) - L(\mathbf{x}_i)|}{\mathcal{K}}, \\ \frac{\sqrt{\sum\limits_{i}^{\mathcal{K}} (f(\mathbf{x}_i) - L(\mathbf{x}_i))^2}}{\mathcal{K}}, \\ \max_{i \in \mathcal{K}} |f(\mathbf{x}_i) - L(\mathbf{x}_i)|, \end{gathered} \tag{10}$$

where $\mathcal{K}$ 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 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 the rules (11) performs the mapping $L : \Re^{n+m} \rightarrow \Re$ 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:

$$\begin{aligned} &L(\mathbf{x}) = \\ &= \frac{\sum\limits_{i=1}^{q} \mu_{LX_{1i}}(x_1) \cdot \mu_{LX_{2i}}(x_2) \cdot \ldots \cdot \mu_{LX_{(n+m)i}}(x_{n+m}) \cdot r_i}{\sum\limits_{i=1}^{q} \mu_{LX_{1i}}(x_1) \cdot \mu_{LX_{2i}}(x_2) \cdot \ldots \cdot \mu_{LX_{(n+m)i}}(x_{n+m})}, \end{aligned} \tag{11}$$

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 folows:

$$\begin{aligned} &L(\mathbf{x}) = \\ &= \frac{\sum\limits_{i=1}^{q} \mu_{LX_{1i}}(x_1) \cdot \ldots \cdot \mu_{LX_{(n+m)i}}(x_{n+m}) \cdot z}{\sum\limits_{i=1}^{q} \mu_{LX_{1i}}(x_1) \cdot \mu_{LX_{2i}}(x_2) \cdot \ldots \cdot \mu_{LX_{(n+m)i}}(x_{n+m})}, \end{aligned} \tag{12}$$

where $z = (r_{0i} + r_{1i}x_1 \ldots + r_{(n+m)i}x_{n+m})$.

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 (1) 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, (1) can be represented as a fuzzy Takagi-Sugeno (TS) model [20]. 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 $Y_1^\theta, \ldots, y(t-r)$ is Y_r^θ,
$x(t)$ is $X_0^\theta, \ldots, x(t-s)$ is X_r^θ, then

$$\begin{gathered} y^\theta(t) = a_0^\theta + \sum_{k=1}^{r} a_k^\theta y(t-k) + \sum_{l=0}^{s} b_l^\theta x(t-l), \\ \theta = 1, \ldots, n, \end{gathered} \tag{13}$$

where $a^\theta = (a_0^\theta, a_1^\theta, \ldots, a_r^\theta), b^\theta = (b_0^\theta, b_1^\theta, \ldots, b_s^\theta)$ are adjustable parameter vectors; $\mathbf{y}(t-r) = (1, y(t-1), \ldots, y(t-r))$ is state vector; $\mathbf{x}(t-s) = (x(t), x(t-1), \ldots, x(t-s))$ 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:

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

where $m = r + s + 1$, one obtains the analytic form of the fuzzy model (4), intended for calculating the output $\hat{y}(t)$:

$$\widehat{y}(t) = c^T \widetilde{\boldsymbol{u}}(t), \tag{14}$$

where $c = (c_0^1, \ldots, c_0^n, \ldots, c_m^1, \ldots, c_m^n)^T$ is the vector of the adjustable parameters;

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

is the extended input vector;

$$\beta^\theta(t) = \frac{U_1^\theta(u_1(t)) \otimes \ldots \otimes U_m^\theta(u_m(t))}{\sum_{\theta=1}^{N} \left(U_1^\theta(u_1(t)) \otimes \ldots \otimes U_m^\theta(u_m(t))\right)} \tag{15}$$

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 $\boldsymbol{c}(t)$ is calculated using the known multi-step LSM:

$$\boldsymbol{c}(t) = \boldsymbol{c}(t-1) + Q(t)\widetilde{u}(t)[y(t) - \boldsymbol{c}^T(t-1)\widetilde{\boldsymbol{u}}(t)] \tag{16}$$

$Q(0) = \gamma I$, $\gamma \gg 1$ 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:

$$\begin{matrix} \beta_1^{\Theta_t} & \ldots & \beta_P^{\Theta_t} \\ \ldots & \ldots & \ldots \\ \beta_1^{\Theta_{t-s}} & \ldots & \beta_P^{\Theta_{t-s}} \end{matrix} \tag{17}$$

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 the condition (4) will determine the predicate Ξ 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 selection combined with the 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.

Fuzzy virtual 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:

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

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.

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 [21].

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 $l = \{\overline{x}_N, \ldots \overline{x}_1\}$. 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 $\overline{x}_i$, $i = 1, 2, \ldots N$ can be described by means of a set of variables ("characteristics") $x_{i1}, x_{i2}, \ldots x_{iS}$. The set $\{x_{i1}, x_{i2}, \ldots x_{iS}\}$ 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 $\forall i = 1, \ldots, N$. As $x_i = x_{i1}, x_{i2}, \ldots x_{iS}$, we denote a set of observations for the object $\overline{x}_i, i = 1, 2, \ldots N$. The set of variables' observations relevant to the sample will be represented as a data table V with N rows and S columns: $V = \{x_{ij}\}$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, S$; 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.

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:

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

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

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

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

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

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

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

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

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

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:

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

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).

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 where the control plant is nonlinear, high response speed is not required, and the computational resources enable the exhaustive search of the process history. Such approach is quite satisfactory for the identification of a rather broad class of control plants, foe example, continuous and semibatch 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 picking 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.

Application of traditional clusterization methods for 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 $\overline{x}_i \in \mathbf{X}$, $i = 1, \ldots, N$ is split into several disjoint subsets. Here, any object from $\mathbf{X}$ 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.

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

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

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

$\varnothing \subset \mathbf{A}_k \subset X$, $k = 1, \ldots, K$ – 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:

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

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

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

The matrix $\mathbf{U}$ should have the following properties:

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

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

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:

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

where $\overline{x}_i \in \mathbf{X}$, $i = 1, \ldots, N$ is the i-th object of clustering, $\mathbf{A}_k = \|\overline{x}_p\|$, $u_{kp} = 1$, $p = 1, \ldots, N$ is the cluster with the number k, $g_k = \dfrac{1}{|\mathbf{A}_k|} \sum\limits_{x_i \in \mathbf{A}_k} \|x_i\|$ is the center of the k-th cluster.

2. 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 $\mathbf{F} = [\mu_{ki}]$, $k = 1, \ldots, k$; $i = 1, \ldots, N$. Typically, $\mu_{ki} \in [0, 1]$ are selected, where the k-th row of the matrix $\mathbf{F}$ denotes that the objects $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_N$ belong to the cluster $\mathbf{A}_k$, $k = 1, \ldots, K$.

The only difference between the matrices $\mathbf{F}$ and $\mathbf{U}$ 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:

$$\sum_{k=1}^{K} \mu_{ki} = 1, \qquad i = 1, \ldots, N,$$

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

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 (Pascual-Marqui, 2001):

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

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

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

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

Solving the associative search 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:

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

where $\overline{x}_N \in \mathbf{X}$ 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

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

is attained for $k = r$. Further, $\overline{x}_N$ 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 $\overline{x}_N$. In such case not only a considerable number of vectors distant from $\overline{x}_N$ 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 $\overline{x}_N$ 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.

Applications

The presented methods were successfully applied in soft sensor design for chemical and oil refining processes. The approach proposed in [17] is based on virtual models and associative search techniques. A fuzzy model is applied in combination with production knowledgebase to compensate for the lack of lab data.

For short-term prediction in trading identification algorithms based on associative search procedure can be used [18]. In case of fuzzy definition of one or more variables, successful associative search is possible only with knowledgebase built and augmented by analyst or trader during a trading session.

The paper [19] presents a soft sensor describing the dynamic behavior of power grid's generation facilities participation in the overall primary frequency regulation in contingencies. The soft sensor is based on generating capacity and frequency time series and was developed by the authors for the Control Center of Russia's Unified Energy System (RAO UES) of Russia.

The establishment of a power plant's participation and the estimation of the degree of its contribution to the overall primary frequency regulation (OPFR) is performed at the frequencies exceeding 0.2 Hz. When the grid is operated in design mode (with frequency deviations less than 0.2 Hz), the control is purely qualitative and informative.

At the same time, the qualitative assessment of generating facilities' participation in the OPFR at abrupt frequency excursions in the grid in the range 0.05…0.2 Hz, the systematic (more than 50% cases over a year) nonparticipation in the primary regulation of the generating facilities from a number of heat power plants was detected due to the lack of the requisite power adjustments to compensate for the frequency deviations.

Along with the establishment the fact of specific generating facility's participation (or nonparticipation) in the OPFR, the technology of RAO UES Control Center's historical data processing enables the evaluation of the degree of plant's participation in the regulation.

To rank the generating facilities' participation in the OPFR, identification models and algorithms describing power grid's dynamics were developed.

When building the identification model, essential non-linearity of the object under investigation was allowed for. Therefore, it was found rational to use associative search models in the soft sensor design.

The identification algorithms show the generation facilities ranking w.r.t. the probabilities of their violation of the OPFR participation requirements to the generating facilities. The indicators of specific generating facilities influence on the OPFR were evaluated that contributed to the quality improvement of the secondary frequency regulation.

An intelligent system intended for dynamic state estimation of a complex power grid was created per the EU project "**I**NTELLIGENT **C**OORDINATION OF **O**PERATION AND EMERGENCY CONTROL OF **EU** AND **R**USSIAN POWER GRIDS" (***ICOEUR***, FP7-ENERGY) will be running from 2007 to 2013. The system is underpinned by intelligent algorithms of grid dynamics identification with automatic on-line self-tuning based on the data from monitoring systems.

State estimation models for power facilities with on-line model tuning are based on data monitoring and application of a new predictive method for state estimation – the associative search method.

The acquisition, storage, processing, displaying, analysis and documenting of the information are executed in real time based on the data from automated power generation, distribution and consumption systems and supervisory control, monitoring and accounting.

The development of intelligent dynamic state estimation algorithms based on the use of process knowledge for important power plant and power network control tasks such as the optimization of generating equipment and power grid optimization is in sight.

Fuzzy models will underlie those; virtual object models using the associative search method will be also employed.

State estimation models, customizable during real-time operation, can be used both in the automatic mode of a control system, and to support management decisions.

Based on the dynamically configurable state models, power facility operation modes can be optimized over the whole of power grid including all power market participants subject to reliability and profitability criteria.

Predicting the qualities of delayed coking unit distillates

Quality control of delayed coker distillates for their subsequent utilization as hydrotreater's feedstock is a challenge. The reason is that coker's fractionator was originally designed for some average feed rate and quality, while in real life both change several times per day sufficiently for making a serious disturbance for downstream process equipment.

The traditional control strategy for a fractionator under disturbances is temperature profile stabilization closer to the steady-state values established at design phase for average feed rate and quality, and their further slight adjustment subject to lab data. The product samples are analyzed by refinery lab 3–6 times per day. This makes the control strategy ineffective because the object's state cannot be identified unambiguously from such scanty samples.

Soft sensors (SS)-based virtual models were built for this plant using both process history and lab data. Those models enabled on-line calculation of desirable product qualities with sufficient accuracy. This resulted in process unit's throughput increase combined with more consistent product quality.

The SS-based quality analyzers were built for coker naphtha IBP, 50%, 90% distillation points, and EP and coker gas oil IBP, 50%, and 90% distillation points. Based on these, a predictive model structure for distillation points of key product streams was obtained as well as the forecast accuracy estimate.

The forecast was calculated using a mathematical model whose inputs were process variable measured on-line. The forecast accuracy depends on right selection of informative variables, memory depth, and the amount of available plant data.

Typically, the precise forecast is impossible for such complex objects as fractionator because the existing measurements do not observe all factors affecting the product qualities. For example, there were no tools at the process unit to measure feedstock make-up changes. In such case, the informative variables had to be selected from the vast amount of data. This was done using process history. At design operation of the model-based predictor, its adaptation to plant dynamics and input properties changes is executed automatically with the changes of the nonlinear model's structure, while its dynamic depth remains the same.

Figure 3 shows an example of a predictive model for coker naphtha 50% distillation point (ASTM D86).

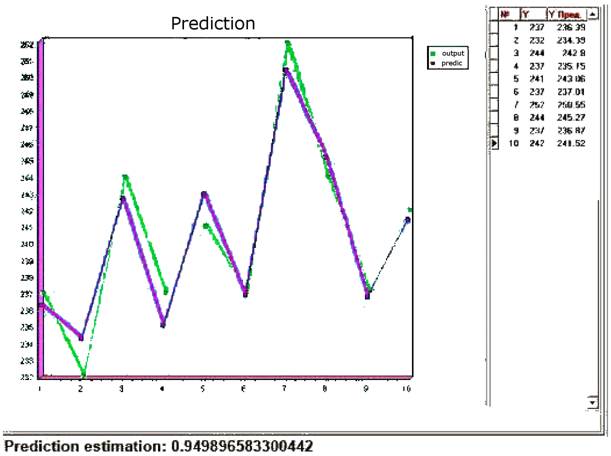

Fig. 3. Naphtha 50% distillation point model.

The SS-based control system can calculate control actions with adaptive models adjustment.

After site acceptance tests in advising mode are complete, the recommended control action can be further used in the closed loop, i.e., in the automatic control system with an identifier (Fig. 4).

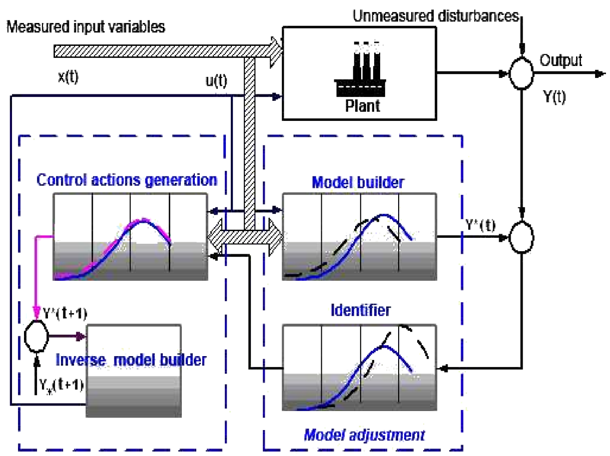

Fig. 4. Adaptive control system with identifier.

Conclusions

To develop decision-making support for process operator, soft sensors are offered underlain by a novel approach to identification. The approach presumes forecast generation on every time step based on *virtual models* rather than the time approximation of the process. To simplify the identification algorithm, an *associative search* procedure is offered based on process knowledge employment for generating the images of the variables under investigation. The application of this approach is especially effective whilst compensating for insufficient lab data for model development. In such case, fuzzy specification of certain process variables using process knowledgebase is practiced.

References

[1] Bellman R. Kalaba R., *Dynamic Programming and Modern Control Theory*, Academic Press, New York, N.Y., 1965.

[2] Bunich A., Lototsky V., *Fast convergent identification algorithm for limited noise*, Proc. of 15th IFAC World Congress, Barcelona, Spain, 2002.

[3] Goodwin G., Ramadge P., Caines P., *Discrete time stochastic adaptive control*, SIAM Journal on Control and Optimization, 19, 829–853, 1981.

[4] Juditsky A., Nazin A., *On minimax approach to nonparametric adaptive control*, Int. J. Adapt. Control & Signal Process, 15, 153–168, 2001.

[5] Ljung L., *System identification. Theory for user*, Prentice Hall, Upper Saddle River, N.J., 2nd edition, p. 609, 1999.

[6] Poznyak A.S., Kaddour N., *Learning Automata and Stochastic Optimization*, Springer Verlag, New York, 1997.

[7] Chen H.F., *Recursive system identification and adaptive control by use of the modified least squares algorithm*, SIAM J. Control and Optimization, 22, 5, 758–776, 1984.

[8] Ljung L., *Perspectives on System Identification*, Plenary on 17th IFAC World Congress, Seoul, Korea, 2008.

[9] Larichev O.I., Asanov A., Naryzhny Y., Strahov S., *Expert System for the Diagnostics of Acute Drug Poisonings, Applications and Innovations in Intelligent Systems IX*, Proc. of the 21 SGES International Conference on Knowledge Based Systems and Applied Artificial Intelligence. Cambridge, UK: Springer-Verlag, 159—168, 2001.

[10] Patel V.L., Ramoni M.F., *Cognitive Models of Directional Inference in Expert Medical Reasoning* in Feltovich P., Ford K., Hofman R. (Eds.) Expertise in Context: Human and Machine. AAAI Press, Menlo Parc, CA., 1997.

[11] Hunt E., *Cognitive Science: Definition, Status and Questions,* Annual Review of Psychology, 40, 603–629, 1989.

[12] Simon H.A., *The Sciences of the Artificial* (3rd ed.), The MIT Press, Cambridge, MA, 1997.

[13] Costa E.F., Tinos R., Oliveira V.A., Araujo A.F.R., *Reinforcement learning schemes for control design*, Proc. of American Control Conference, 4, 2414–2418, 1997.

[14] Gavrilov A.V., *The model of associative memory of intelligent system*, Preprints of the 6-th Russian-Korean International Symposium on Science and Technology, 1, 174–177, 2002.

[15] Larichev O. Brown R., *Numerical and verbal decision analysis: Comparison on practical cases*, Journal of Multi-Criteria Decision Analysis, 9, 263–273, 2000.

[16] Chadeev V.M., *Digital Identification of Nonlinear Dynamic Plants*, Automation and Remote Control, 12, 85–93, 2004.

[17] Bakhtadze N., Lototsky V., Maximov E., Pavlov B., *Associative search models in industrial systems*, Proc. of IFAC workshop on intelligent manufacturing systems, 156–161, 2007.

[18] Bakhtadze N., Lototsky V., Valiakhmetov R. *Associative search models in trading*, Proc. of 17th IFAC World Congress, Seoul, Korea, 4280–4284, 2008.

[19] Bakhtadze N., Maximov E., Valiakhmetov R.E., *Fuzzy soft sensors for chemical and oil refining processes*, Proc. of 17 IFAC World Congress, Seul, Korea, 4246–4250, 2008.

[20] Takagi T., Sugeno, M., *Fuzzy identification of systems and its applications to modeling and control*, IEEE Trans. Systems Man Cybernetics, 15 (1), 116–132, 1985.

[21] Jain A.K., Murty M.N., Flynn P.J., *Data clustering: a review*, ACM Computing Surveys, 31 (3), 264–323, 1999.

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