RECURRENT PROCEDURE IN SOLVING THE GROUPING INFORMATION PROBLEM IN APPLIED MATHEMATICS

V. Donchenko, Yu. Krivonos, Yu. Krak

***Abstract**: Number of The grouping information problem in its two basic manifestations recovering function, represented by empirical data (observations) and problem of classification (clusterization) and conception of its solving by the standard recurrent procedures are proposed and discussed. It is turn out that in both case correspond procedures can be designed on the base of so called neurofunctional transformations (NfT—transformations). Conception of such transformations implements the idea of superposition of standard functions by certain sequence of recurrent applications of the superposition. Least Square Method is used for designing the elementary functional transformations and implemented by necessary developed of M-P-inverse technique. It is turn out that the same approach may be designed and implemented for solving the classification problem. Besides, the special classes of beam dynamics with delay were introduced and investigated to get classical results regarding gradients. These results were applied to optimize the NfT—transformations.*

***Keywords**: Grouping information problem, generalized artificial neuronets, learning samples, beam dynamics, Fuzzy likelihood equation, Multiset theory.*

***ACM Classification Keywords**: G.2.m. Discrete mathematics: miscellaneous,G.2.1 Combinatorics. G.3 Probability and statistics, G.1.6. Numerical analysis I.5.1.Pattern Recognition H.1.m. Models and Principles: miscellaneous:*

Introduction

The problem of grouping the information (grouping problem) is the fundamental problem of applied investigations. It appears in various forms and manifestations. All of them eventually are reduced to two forms. Namely, these are: the problem of recovering the function represented by their observations and the problem of clustering, classification and pattern recognition. State of art in the field is represented perfectly in[Kohonen,2001], [Vapnik, 1998], [Haykin, 2001], [Friedman, Kandel,2000], [Berry,2004] .

It's opportune to mark what the information regarding the object or a collection of similar object is exposed to aggregating is. It is of principal importance that an object is considered as a set of its main components and fundamental for the object ties between them. Such consideration and only this one enable application of the math in object description, namely, for math modeling. It is due the fact that after Georg Cantor the objects of investigation in math (math structures) are the sets plus "ties" between its elements. There are only four (may be, five) fundamental mathematical means to describe these "ties". Namely, these are: relations, operations, functions and collections of subsets (or combinations of mentioned above). Thus, the mathematical description of the object (mathematical modeling) cannot be anything other than representing the object structure by the means of mathematical structuring. It is applicable to the full extent to that objects which indicated by the term “complex system”. A “complex system” should be understanding and, correspondingly, determined, as an objects with complex structure (complex “ties”). Namely, when reading attentively manuals by the theme (see, for example, [Yeates, Wakefield, 2004], [Forster , Hölzl, 2004]) one could find correspondent allusions. It is reasonable understanding of “complex systems” instead of the its understanding as the “objects, consisting of numerous parts, functioning as an organic whole”.

So, math modeling is designing in math “parts plus ties”, which reproduce “part plus ties” in reality.

So it is principal question in math modeling which math objects represents “part” of the object and which the “ties” ones. The math object - representative should be chosen in such a way that variety of math structuring means were sufficient to convey the object structure.

It is commonly used approach for designing objects - representative to construct them as an finite ordered collection of characteristics: quantitative (numerical) or qualitative (non numerical). Such ordered collection of characteristics is determined by term cortege in math. Cortege is called vector when its components are numerical. In the function recovering problem objects - representatives are vectors and functions are used as a rule to design correspond mathematical “ties”. In clustering and classification problem the collection may be both qualitative and quantitative. In last case correspond collection is called feature vector. It is reasonable to note that term “vector” means more, than simply ordered numerical collection. It means that curtain standard math “ties” are applicable to them. These “ties” are adjectives of the math structure called Euclidean space denoted be R^n. Namely these are: linear operations (addition and scalar multiplying), scalar product and correspond norm.

Just the belonging to the base math structure (Euclidean space) determines advantages of the “vectors” against “corteges”. It is noteworthy to say, that this variant of Euclidean space is not unique: the space $R^{m\times n}$ of all matrixes of a fixed dimension $m\times n$ may represent alternative example. The choice of the R^n space as “environmental” structure is determined by perfect technique developed for manipulation with vectors. These include classical matrix methods and classical linear algebra methods. SVD-technique and methods of Generalized or Pseudo Inverse according Moore – Penrose are comparatively new elements of linear matrix algebra technique [Nashed, 1978](see, also, [Albert,1972] , [Ben-Israel, Greville ,2002]). Outstanding impacts and achievements in this area are due to N.F Kirichenko (especially, [Кириченко, 1997] [Kirichenko, 1997], see also [Кириченко, Лепеха, 2002]). Greville’s formulas:forward and inverse -for pseudo inverse matrixes, formulas of analytical representation for disturbances of pseudo inverse, - are among them. Additional results in the theme as to furter development of the technique and correspondent applications one can find in [Кириченко, Лепеха, 2001], [Donchenko , Kirichenko,Serbaev, 2004], [Кириченко, Крак, Полищук,2004] [Kirichenko, Donchenko, Serbaev,2005], [Кириченко, Донченко,2005] [Donchenko, Kirichenko , Krivonos, 2007], [Кириченко, Донченко,2007] ,[Кириченко, Кривонос, Лепеха 2007], [Кириченко, Донченко, Кривонос, Крак, Куляс, 2009].

As to technique designing for the Euclidean space $R^{m\times n}$ as “environmental” one see, for example [Донченко, 2011]. Speech recognition with the spectrograms as the representative and the images in the problem of image recognition are the natural application area for the correspond technique.

As to the choice of the collection (design of cortege or vector) it is necessary to note, that good “feature” selection (components for feature vector or cortege or an arguments for correspond functions) determines largely the efficiency of the problem solution.

As noted above, the efficiency of problem solving group, the choice of representatives of right: space arguments or values of functions and suitable families past or range of convenient features vectors. This phase in solving the grouping information problem must be a special step of the correspondent algorithm. Experience showed the effectiveness of recurrent procedures in passing through selection features step. For correspond examples see, [Ivachnenko,1969] with Ivachnenko's GMDH (Group Method Data Handling), [Vapnik, 1998] with Vapnik's Support Vector Machine. Further development of the recurrent technique one may find in Donchenko, Kirichenko,Serbaev, 2004], [Кириченко, Крак, Полищук,2004] [Kirichenko, Donchenko , Serbaev,2005], [Кириченко, Донченко,2005] [Donchenko, Kirichenko , Krivonos, 2007], [Кириченко, Донченко,2007] ,[Кириченко, Кривонос, Лепеха 2007]. The idea of nonlinear recursive regressive transformations (generalized neuron nets or neurofunctional transformations) due to Professor N.F Kirichenko is represented in the works referred earlier in its development. Correspondent technique has been designed in this works separately for each of two its basic form f the grouping information problem. The united form of the grouping problem solution is represented here in further consideration. The fundamental basis of the recursive neurofunctional technique include the development of pseudo inverse theory in the publications mentioned earlier first of all due to Professor N.F. Kirichenko and his disciples.

The essence of the idea mentioned above is thorough choice of the primary collection and changing it if necessary by standard recursive procedure. Each step of the procedure include detecting of insignificant components, excluding or purposeful its changing, control of efficiency of changes has been made. Correspondingly, the means for implementing the correspondent operations of the step must be designed. Methods of neurofunctional transformation (NfT) (generalized neural nets, nonlinear recursive regressive transformation: [Donchenko , Kirichenko,Serbaev, 2004] [Кириченко, Крак, Полищук, 2004], [Кириченко, Донченко, Сербаєв, 2005]).

Neurofunctional transformation in recovering function problem

The fundament of the Math truth is the conception of deducibility. It means that the status of truth (proved statement) has the statement which is terminal in the specially constructed sequence of statements, which called its proof. The peculiarity in sequence constructing means, that a next one in it produced by previous by special admissible rules (deduction rules) from initial admissible statements (axioms and premises of a theorem). As a rule, corresponded admissible statements have the form of equations with the formulas in both its sides. So, each next statement in the sequence-proof of the terminal statement is produced by previous member of sequence (equation) by changing some part of formulas in left or right it side on another: from another side of equations-axioms or equations premises. The specification of the restrictions on admissible statements and the deduction rules are the object of math logic.

As it was already marked, the idea of neurofunctional transformation (NfT-) or neurofunctional transformation in recovering function problem in the variant of inverse recursion was offered in [Кириченко, Крак, Полищук, 2004], and in variant of forward recursion - in [Donchenko , Kirichenko,Serbaev, 2004] , [Кириченко, Донченко, Сербаєв, 2005]. References on neuronets is determined by the fact that NfT generalizes artificial neuronets: in possibilities of the standard functional elements (ERRT(elementary recursive regression transformation) in NfT): in topology of its connection; in adaptive design of NfT structure in the whole; in adequate math for its description. Just this forward variant will considered below. Namely , NfT- is the transformation built by recursive application of the certain standard element, which will be designated by abbreviation ERRT (Elementary Recursive

Regression Transformer). Process of construction of the NfT- transformation consists in connection of the next ERRT (or certain number of it) to already constructed during previous steps transformer according to one of three possible types of connection (connection topology). Types of connection which will be designated as “parinput”, “paroutput” and “seq”, realize natural variants of use of an input signal: parallel or sequential over input, - and parallel over output. An input of the Output of current step of recursion is input of the next step.

The basic structural element of the NfT- -transformer is ERRT - an element [Кириченко, Донченко, Сербаєв, 2005], which is determined as mapping from R^{n-1} in R^m of a kind:

$$y = A_+ \Psi_u \left(C \begin{pmatrix} x \\ 1 \end{pmatrix} \right) \tag{1}$$

which approximates the dependence represented by training sample

$$(x_1^{(0)}, y_1^0), ..., (x_M^{(0)}, y_M^{(0)}), x_i^{(0)} \in R^{n-1}, \ y_i^{(0)} \in R^m, \ i = \overline{1,M},$$

where:

- C–(n×n) - matrix, which performs affine transformation of the vector $x \in R^{n-1}$ - an input of the system; it is considered to be given at the stage of synthesis of ERRT;
- Ψ_u – nonlinear mapping from R^n in R^n, which consists in component-wise application of scalar functions of scalar argument $u_i \in \Im$, $i = \overline{1,n}$ from the given final set $\Im$ of allowable transformations, including identical transformation: must be selected to minimize residual between input and output on training sample during synthesis of ERRT;
- A+– solution A with minimal trace norm of the matrix equation

$$AX_{\Psi_u C} = Y, \tag{2}$$

in which matrix $X_{\Psi_u C}$ formed from vector-columns $\Psi_u(C \begin{pmatrix} x_i^{(0)} \\ 1 \end{pmatrix}) = \Psi_u(z_i^{(0)})$, and Y – from columns, $y_i^{(0)}, i = \overline{1,M}$.

In effect, ERRT represents empirical regression for linear regression y on $\Psi_u \left(C \begin{pmatrix} x \\ 1 \end{pmatrix} \right)$), constructed with method of the least squares, with previous affine transformation of system of coordinates for vector regressor x and following nonlinear transformation of each received coordinate separately.

Remark 1. Further we shall assume that functions of component-wise transformations from $\Im$ would have a necessary degree of smoothness where it is necessary.

Task of synthesis of ERRT by an optimal selection of nonlinear transformations of coordinates on the given training sample was introduced and solved in already quoted above work [Кириченко, Донченко, Сербаев, 2005]. The solution of a task of synthesis is based on methods of the analysis and synthesis of the pseudoinverse matrices, developed in [Кириченко, 1997]. Particularly, reversion of Grevil’s formula [10] was proved in these works, that recurrently allows to recalculate pseudoinverse matrices when a column or a row of the matrix changed by another one.

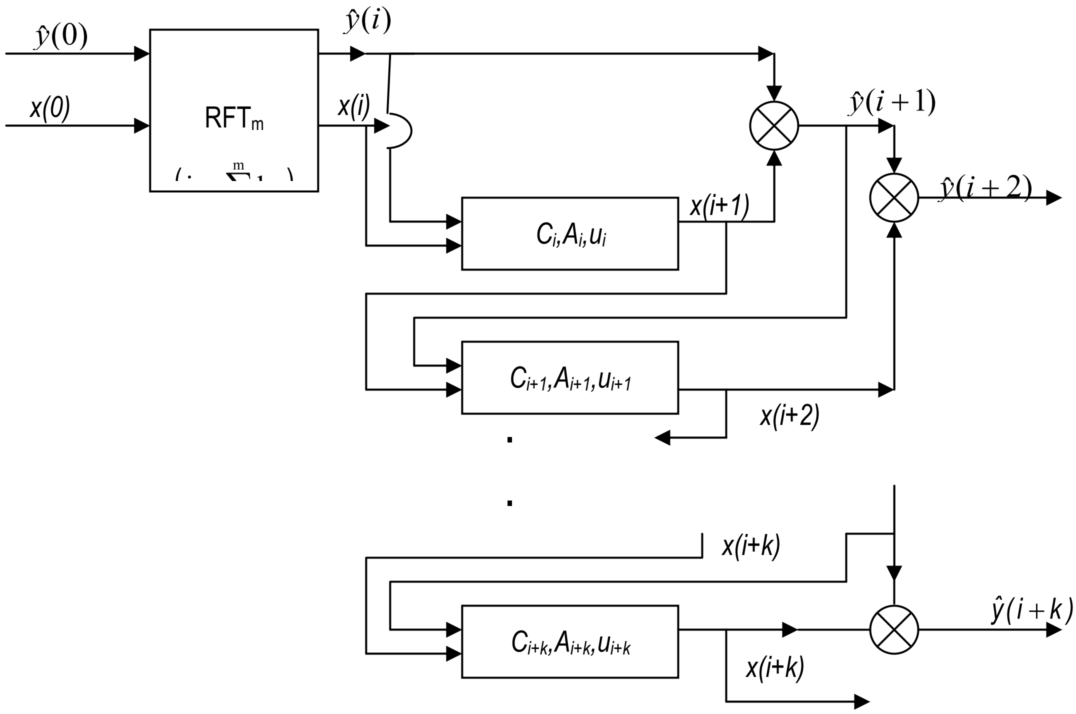
Recurrent procedure in NfT- design: topology and mathematics

Recursion in construction of the NfT--transformer in variant of forward recursion offered below will be considered in the generalized variant in which several ERRTs is used in recurrent connection to already available NfT--structure. Total quantity of recurrent references we shall designate through N, and quantity of ERRTs used on a step m – by $k_m, m=\overline{1,N}$. The common number of ERRTs, used for construction of whole transformer will be designated by $T : T=\sum_{m=1}^{N} k_m$.

Variants of the generalized forward recursion which depend from type of connection of attached ERRT: parinput, paroutput and seq, - are represented on figures 1-3. Where $\hat{y}$ designates an output of already available NfT--structure or an output of the same structure after connection of the next ERRT from the total number k_m of such elements, attached on a step m of the recursion: $m=\overline{1,N}$. Each figure is accompanied by the system of equations which determine transformation of a signal on the next step of recursion.

Type parinput:

Figure 1. Scheme of connection of parinput type – forward recursion.



In parinput type of connection already constructed structure approximates an output of training sample by its input, and set of ERRTs - resulting residual of such approximation depending from an input of training sample. Transformation of the information at this type of connection is described by the following system:

$$\begin{aligned} &x(i+j)=A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}x(i)),\\ &\hat{y}(i+j)=\hat{y}(i+j-1)+A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}\cdot x(i))\\ &i=\sum_{l=1}^{m}k_l,\ j=\overline{1,k_{m+1}} \end{aligned} \tag{3}$$

Type paroutput:

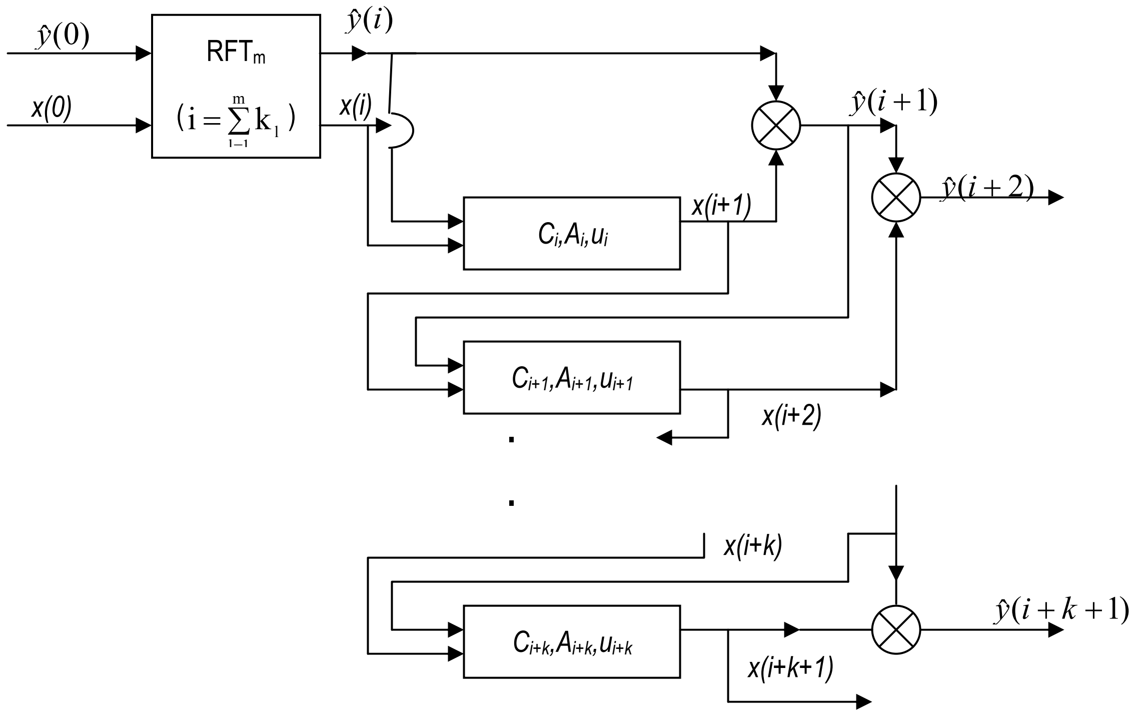


Figure 2. Scheme of connection of paroutput type – forward recursion.

The system describing transformation inside the transformer and communication between an input and an output, at this type of connection looks like:

$$\begin{aligned} & x(i+j) = A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}x(i+j-1)), \\ & \hat{y}(i+j) = \hat{y}(i+j-1) + A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}\cdot x(i+j-1)) \\ & i = \sum_{l=1}^{m} k_l,\ j = \overline{1, k_{m+1}} \end{aligned} \tag{4}$$

Type seq:

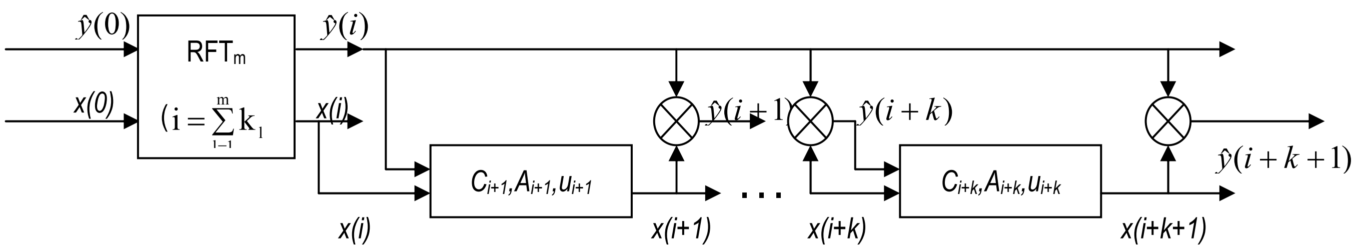


Figure 3. Scheme of connection of seq type – forward recursion.

The equations describing transformation of the information on the next step of recursion look as follows:

$$x(i+j) = A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}x(i+j-1)),$$
$$\hat{y}(i+j) = \hat{y}(i) + A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}\cdot x(i+j-1)) \qquad (5)$$
$$i = \sum_{l=1}^{m} k_l,\ j = \overline{1,k_{m+1}}$$

In this scheme of connection RFT_{m-1} approximates an output part of training sample by input part, and set of ERRTs approximates residual which depends from an output of the next ERRTs.

Entry conditions for all types of connections are described by equations:

$x(0) = x$ – an input of whole NfT--transformer, (6)

$\hat{y}(0) = 0, \hat{y}(1) = x(1)$ for all types of connections. (7)

According to (6) in a training mode the inputs of NfT--transformer are $x_{i1}^{(0)} : x_i^{(0)} \in R^{n-1}, y_i^{(0)} \in R^m, i = \overline{1,M}$ and outputs are – $y_i^{(0)} \in R^m, i = \overline{1,M}$.

Connections of recursive construction of the NFT--transformer are determined so, that standard functional of the least squares method is minimized during its construction, i.e.

$$\sum_{i=1}^{M} \| y_i^{(0)} - RFT(x_i^{(0)}) \|^2 \qquad (8)$$

Equations (3) - (7) for N steps of recursion with common number T of used ERRTs, $T = \sum_{m=1}^{N} k_m, k_m$,– number of ERRTs, used on a step with number $m = \overline{1,N}$, and also efficiency functional (8) - represent mathematical model of NfT- transformation.

Math for NfT: Discrete control systems with delay

Equations (3)-(8) represent the system of the recurrent equations being certain generalization of a classical control system with discrete time (for details see, for example [Бублик, Кириченко,1975]) and first of all in referring to delay.

The simple and combined control systems with delay

Definition. Simple, accordingly - combined, - nonlinear control system with delay on a time interval $\overline{0,N}$ is a control system whose trajectories are defined by system of recurrent equations (9), accordingly - (10), entry conditions (11) and efficiency functional (12) and represented below:

$$x(j+1) = f(x(j-s(j)), u(j), j) \ , \qquad (9)$$

$$x(j+1) = f(x(j), x(j-s(j)), u(j), j) \ . \qquad (10)$$

$$j = \overline{0,N-1}, \qquad x(0) = x^{(0)}, \qquad (11)$$

$$I(U) = \Phi(x(\mathrm{N})), \qquad (12)$$

where function s(j), k=0,...,N–1: s(0)=0, $s(j) \in \{2,...,j\}, j = \overline{0,N-1}, j = \overline{0,N-1}$ – is known.

Evidently, systems with delay for a beam of trajectories are determined. Entry conditions of trajectories of a beam we shall designate $(x(0))_i = x_i^{(0)}, i=\overline{0,M}$, M– number of trajectories of a beam. Trajectories of a beam for both types of systems with delay we shall designate by the appropriate indexation: $x_i(j), j=\overline{0,N-1}, j=\overline{0,M}$,

Let's define efficiency functional for a beam of dynamics which we shall consider dependent only from final states of trajectories of a beam, with equation:]

$$I_p(U)=\sum_{i=1}^{M}\Phi^{(i)}(x_i(N)). \tag{13}$$

Remark 2. Evidently, efficiency functional, as well as for classical control systems, may be defined on all trajectory or trajectories. However, systems with delay at which efficiency functional depends only from final states of a trajectory will be considered in context of NfT--transformers.

Phase trajectories of simple systems with delay are defined by a set of functions f(z, u, j), $j=\overline{0,N-1}$ with one argument z, responsible for a phase variable, and for combined one - a set f(z,v,u,j), $j=\overline{0,N-1}$ with two variables: z, v, which respond for phase variables. Gradients on a phase variables will be denoted by $grad_v f$.

The problem of optimization for both types of such systems with delay is being solved, as well as in a classical case, with construction of the conjugate systems and functions of Hamilton. The assumptions of the smoothness providing correct construction of conjugate systems and functions of Hamilton, and also their use for gradients calculations on controls completely coincide with classical and further will be considered automatically executed.

Optimization in simple control systems with delay

As the analysis of the numerous sources on Theory of Probability and Math Statistics [Донченко 2009], notion of experiment in them is associated with something, named conditions (condition of experiment), under which phenomena is investigated, and something, that appears under the conditions: named the results of experiment.

So, as in [Донченко 2009] "experiment" is proposed to be considered the pair *(c, y)*: *c*- conditions of experiment (observation, trail, test), y – result of experiment. Henceforth Y_c for the fixed condition *c* will denote the set of all possible that may appear in the experiments under conditions $c \in C$. Generally speaking Y_c is not singleton.

It is reasonable to mark out in a condition c variational, controlled, part x: $x \in R^p$ as a rule, and part f, which is invariable by default in a sequence of experiment. Condition c under such approach is denoted be the pair: c=(x, f), $x \in X \subseteq R^p$.

Definition. The conjugate system of a simple control system with delay (9) - (13) we shall call following recurrent equation concerning $p(k): k=\overline{N,0}$:

$$p(k)=\sum_{j\in\{j:j-s(j)=k,\geq kj\}} grad_{x(k)}\{p^T(j+1)f(x(j),u(j),j)\} \quad ,$$

$$k=\overline{N-1,0}$$

with the initial condition

$$p(N)=-grad_{x(N)}\Phi(x(N)).$$

Accordingly, in the case of a beam of trajectories the conjugate systems are defined for each trajectory by equations:

$$p^{(i)}(N) = -grad_{x_i(N)}\Phi^{(i)}(x_i(N)),$$

$$p^{(i)}(k) = \sum_{j\in\{j:j-s(j)=k,j\geq k\}} grad_{x_i(k)}\{p^{(i)T}(j+1)f(x_i(j),u(j),j)\} \quad ,$$

$$k=\overline{N-1,0}, i=\overline{1,M}.$$

Function of Hamilton for simple system with delay is defined by a classical equation:

$$= p^T(k+1)f(x(k-s(k)),u(k),k), k=\overline{N-1,0}$$

For a beam of trajectories of a simple control system with delay the set of functions of Hamilton $\mathrm{H}^{(i)}$, $i=\overline{1,M}$ for each of trajectories $x_i(k), k=\overline{N-1,0}$.

Theorem 1. The gradient on control from the efficiency functional which depends only from final states of trajectories of a beam, for a simple control system with delay is defined by equations:

$$grad_{u(k)}I(U) = -\sum_{i=1}^{M} grad_{u(k)}\left\{p^{(i)T}(k+1)f(x_i(k-s(k)),u(k),k)\right\} =$$

$$= -grad_{u(k)}\sum_{i=1}^{M} H^{(i)}(x_i(k),u(k),k),\ k=\overline{N-1,0}$$

The proof. The proof will be carried out precisely the same as in a classical case: first - for one trajectory, and then by use of additivity of efficiency functional on trajectories of system.

Optimization in combined control systems with delay

Definition . The conjugate system for the combined control system with delay is the system determined by recurrent equations:

$$p(k) = grad_z\{p^T(j+1)f(x(k),x(k-s(k)),u(j),j) \quad +$$

$$+ \sum_{j\in\{j:j-s(j)=k,\geq kj\}} grad_v\{p^T(j+1)f(x(k),x(j),u(j),j)\},$$

$$k=\overline{N-1,0}, i=\overline{1,M},$$

with the initial condition

$$p(N) = -grad_{x(N)}\Phi(x(N)).$$

Accordingly, function of Hamilton *H(p(k+1),x(k),x(k–s(k),u(k),k))* of the combined system is defined by a equation:

H(p(k+1),x(k),x(k–s(k),u(k),k))=

$$= H\left(p(k+1,x(k),x(k-s(k)),u(k),k)\right) = p^T(k+1)f(x(k),x(k-s(k)),u(k),k).$$

As before, the upper index $(i): i=\overline{1,M}, p^{(i)}(k), H^{(i)}, k=\overline{N-1,0}$, will define objects for trajectories of a beam.

Theorem 2. Gradients on the appropriate controls from the efficiency functional which depends only from final values of trajectories of the combined control system with delay, are defined by gradients from the appropriate functions of Hamilton:

$grad_{u(k)} I_p(U) = \text{-}grad_{\ u(k)} \mathrm{H}(\mathrm{p(k+1),x(k),x(k\text{-}s(k)),u(k)),k)})$.

And, hence, for a beam of dynamics the appropriate gradient is defined by equation:

$$grad_{u(k)} I_p(U) = -\sum_{i=1}^{M} \mathrm{grad}_{\ u(k)} \mathrm{H}^{(i)}(\mathrm{p}^{(i)}(\mathrm{k}+1),\mathrm{x}^{(i)}(\mathrm{k}),\mathrm{x}^{(i)}(\mathrm{k\text{-}s(k),u(k),k)}) \tag{14}$$

The result may be proved in a standard way for use of the conjugate systems and functions of Hamilton.

NfT and control systems with delays

As it has been already marked NfT--transformation may be represented by a control system with delay. More precisely, the following theorem is valid.

Theorem 3. Regression RFTN-transformer with the direct N-times recursive reference to $k_m, m = \overline{1,N}$ ERRT elements on each of the steps of the recursion is represented by the nonlinear combined beam of dynamics with delay on the set $\overline{0,T} : T = \sum_{m=1}^{N} k_m$:

- with a phase variable $z(t) = \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix}$ with $z_i(t) \in R^m, i = 1,2, t = \overline{0,T}$;
- by the system of the recurrent equations which determines trajectories of a beam:

$$z^{(i)}(t+1) = f(z^{(i)}(t), z^{(i)}(t-k(t)), C_t, t) =$$

$$= \begin{pmatrix} f_1(z^{(i)}(t), z^{(i)}(t-k(t)), C_t, t) \\ f_2(z^{(i)}(t), z^{(i)}(t-k(t)), C_t, t) \end{pmatrix}, \qquad)$$

with $f_1, f_2 \in R^m, = \overline{1, T-1}, i = \overline{1,M}$, , dependent on topology of the NFT--transformer,

- initial condition

$$z^{(i)}(0) = \begin{pmatrix} x_i^{(0)} \\ 0 \end{pmatrix}, i = \overline{1,M} ,$$

where $z^{(i)}(0), i = \overline{1,M}$, initial conditions of trajectories of a beam, and $x_i^{(0)}, i = \overline{1,M}$,– elements of an input of learn sample;

- efficiency functional $I(C), C = (C_1, \ldots, C_T)$, which depends on matrices $C_1, \ldots, C_T$ as on controls:

$$I(C) = \sum_{k=1}^{M} \| y_k^{(0)} - z_2(T) \|^2 ,$$

where $y_k^{(0)}, k = \overline{1,M}$, components of an output of learning sample.

The proof can be found in [Кириченко, Донченко, Сербаев 2005].

The statement proved enables using of methods of optimization of the theory of control for optimization of already constructed RFTN-transformer on residual size depending on matrices $C_0, C_1, ..., C_{T-1}$. This statement also is a subject of the following theorem.

Theorem 4. By presence of continuous derivatives up to the second order inclusive of functions of family $\Im$ RFT - transformation may be optimized by gradient methods with gradients of the efficiency functional on matrices C, as on parameters.

Proof. According to the theorem 3 NFT--transformation may be represented by the combined control system with delay, and according to the theorem 2 gradients on the appropriate controls - parameters of the NfT-- transformation are defined by equation (14). Importance of the given theorem lies in that it gives exhaustive interpretation of Back Propagation algorithm, outlining at the same time borders of the specified method.

Neurofunctional transformation in classification problem

Clusterization and classification problem as the variant grouping problem will be discussed according [Кириченко, Кривонос, Лепеха, 2007] for two classes $\Omega_x(1), \Omega_x(2) \subseteq R^m$ and, correspondingly, with two correspondent learning samples $x(j) \in \Omega_x(1), j \in J_1$, $x(j) \in \Omega_x(2), j \in J_2 : J_1 \cup J_2 = \{1,...,n\}, J_1 \cap J_2 = \varnothing$. The classification problem will be interpreted as the problem of designing function $\varphi : \varphi \in R^m \to R^1$ (discriminate function), which would "Δ – differentiate" classes for some $\Delta > 0$, in the sense, that:

$\varphi(x(j)) \geq \Delta, j \in J_1, \varphi(x(j)) \leq -\Delta, j \in J_2$

We will find correspond φ, Δ for linear case: when $\varphi(x) = a^T x, a = (a_1, ..., , a_m)^T \in R^m$ (Linear Discrimination Problem (LD -problem)):

$$a^T x(j) = y(j) \geq \Delta, j \in J_1 \tag{15}$$

$$a^T x(j) = y(j) \leq -\Delta, j \in J_2 : \tag{16}$$

$$J_1 \cup J_2 = \{1, ..., n\},$$

$$J_1 \cap J_2 = \varnothing.$$

Under matrix denotation:

$$X = \left(x(1) \vdots \ldots \vdots x(n)\right) = \begin{pmatrix} x_{(1)}^T \\ \ldots \\ x_{(m)}^T \end{pmatrix}, x(j) \in R^m, j = \overline{1,n}, x_{(i)} \in R^n, i = \overline{1,m},$$

$$y = (y(1), ..., y(n)^T \in R^n$$

LD-problem (15),(16) one can rewrite in matrix form under

$$X^T a = y : y \in \Omega_y(\Delta) \subseteq R^n, \Delta > 0, \tag{17}$$

$$\Omega_y(\Delta) = \{y \in R^n : y_j > \Delta, j \in J_1, y_j < \Delta, j \in J_2\} \tag{18}$$

Due to (17), (18) LD-problem we can consider as the solving problem for a System of Linear Algebraic Equaitions (SLAE) with plural form of constraint on $y : y \in \Omega_y(\Delta)$.

Due to results of [Kirichenko, 1997], [Кириченко, Лепехаб2002], see also [Кириченко, Донченко,2005], solvability condition for SLAE (17) is of the next form:

$y^T Z(X) y = 0, Z(X) = I_n - X^+ X$

X^+ - Moore-Penrose pseudoinverse (MP-inverse)(see, for example ,[Albert, 1972]).

Thus, the next lemma is valid.

Lemma 1. Solvability SLAE (17) with plural constraint (18) is equivalent to solvability problem for quadratic equation with constraint of the next form:

$$y^T Z(X) y = 0, \Omega_y(\Delta) \subseteq R^n : \Delta > 0. \tag{19}$$

Due to results of the publication cited before previous Lemma, we have that next statement is valid.

Lemma 2. For each a solution y_d of (19) LD-problem solution is determined by relation

$$a = X^{+T} y_d \tag{20}$$

Thus the next theorem is valid due Lemmas 1,2.

Theorem 5. LD-problem (17)-(18) equivalent to solvability quadratic optimization problem for minimization of quadratic form $y^T Z(X) y$ in domain $\Omega_y(\Delta)$:

$$y_* = \arg \min_{y \in \Omega_y(\Delta)} y^T Z(X) y. \tag{21}$$

Insolvability of the optimization problem from Theorem 5 means insolvability LD-problem with the feature vector of the model. So the features need purposeful change. So, criteria for the choice of correspondent components and means for correspondent changes must be available. Just these means may be realized by the correspondent modification of NfT.

Criteria of informative content for the components of feature vector

There are three criteria for detection of informative value (I-value) for the components of feature vector.

By the first of them component with minimal I-value is the one correspond the number, determined by the relation

$$i^* = \arg \min_{i=\overline{1,m}} y_*^T Z(X_i) y_*. \tag{22}$$

Another criterion is the one determined by the degree of independence of the component from all others:

Coordinate i is I-valuable, if

$$x_{(i)}^T q(i) = \begin{bmatrix} 1, \text{feature is I -valuable} \\ \neq 1, \text{feature is I -valuable} \end{bmatrix}, i = \overline{1,m}, \tag{23}$$

where

$X^+ = (q(1), \ldots, q(m))$.

This criterion based on M-P inverse where (22) (see,[Kirichenko,1997]) is criterion of linear independence of i-component from the rest.

The next criterion is the one for detection minimal I-value component from dependent component of feature vector. The number *i** of the correspondent feature vector component is determined by relation (see,[Kirichenko,1997]):

$$i^* = \arg\min_{i=\overline{1,m}} \frac{\left|y_*^T q(i)\right|^2}{\left\|q(i)\right\|^2}. \tag{24}$$

Algorithm of modification for components of feature vector with minimum of informative content

If the component with minimum informative content is detected it must be changed by its modification Other components of feature vector may be used for such modification. Nonlinear transformation ψ of the component by itself or other components may be used for the modification from certain collection, for example from Table 1:

Table 1.Basic nonlinear transformation

ψ	ψ	ψ
$y = \frac{1}{ax+b}$	$y = \frac{a}{x} + b$	$y = \frac{x}{ax+b}$
$y = \frac{1}{ax^2+bx+c}$	$y = \frac{x}{ax^2+bx+c}$	$y = a + \frac{b}{x} + \frac{c}{x^2}$
$y = ax^b$	$y = ab^x$	$y = ae^{bx}$
$y = ae^{-bx^2}$	$y = ax^b e^{cx}$	$y = ae^{bx+cx^2}$
$y = a\sin(bx+c)$	$y = th(ax)$	$y = Arth(ax)$

Algorithms for modification consists of the next steps.

1. Cycle of solution for task (21).
2. Detection the component, say x_s, for modification according to one of the relations (22)-(24).
3. Modification of the detected component standard procedure, which include changing that component by another: for example $\psi(x_s)$, followed by the choice of best $\psi^* \in \Psi$ according to solution of optimization task:

   $\psi^* = \min_{\psi \in \Psi} y_*^T Z(X_{(s,\psi(x_s))}) y_*$,

   where $X_{(s,\psi(x_s))}$ - matrix, which corresponds to new feature vector with feature $\psi(x_s)$ instead of x_s.
4. New cycle of solution for task (21).

Scheme of the algorithm is represented on Figure 1:

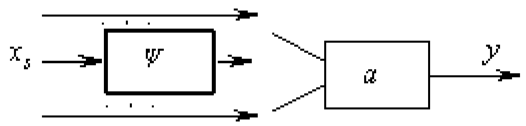


Figure 1.

Modification of step 3 may be of more complicated: with using others of components. For example, next chart (Figure 2) depicts using of nonlinear transformation of another component for changing x_s:

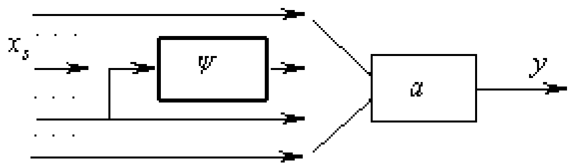


Figure 2.

Some others variants of the modification are represented on the next charts

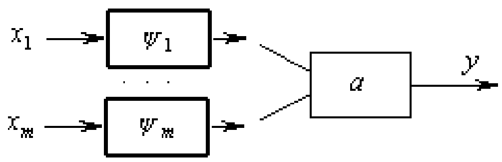


Figure 3.

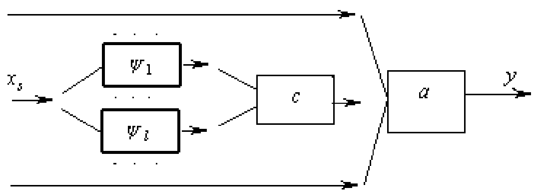


Figure 4.

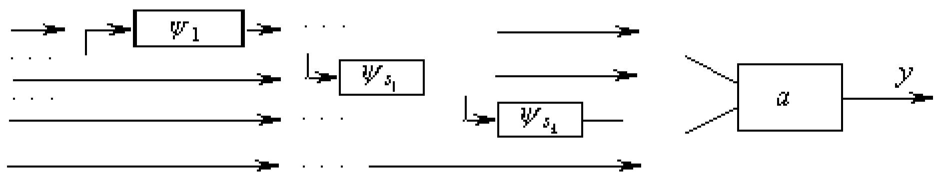


Figure 5.

Chart from Figure 3 represent simultaneous nonlinear transformations for the several component, chart from Figure 4 – modification by changing the component by linear combinations of nonlinear transformations some of others components of feature vector, and Figure 5 represents consequent application of some of one step modification. There may be others variants from those, one can find in [Кириченко, Кривонос, Лепеха, 2007].

Namely charts depicted earlier illustrate the idea of NfT for the transformations of the feature vector.

Conclusion

The realization of the conception for recurrent procedures in solving of important applied grouping information problems is represented. The approach proposed in the article is developed for the both basic form of grouping problem. The development of M-P inverse technique due to Professor Kirichenko and his disciples is the basis for all results of the article.

Bibliography

[Albert ,1972] Albert A.E. Regression and the Moore-Penrose pseudoinverse.-Academic Press.- 1972 .- 180 стор.

[Ben-Israel , Greville ,2002] Adi Ben-Israel ,Thomas N.E. Greville Generalized Inverses .Theory and Applications.- Second Edition.- Springer-Verlag New York, Inc. -2003.-420 p

[Berry,2004] Michael W. Berry Editor Survey of Text Mining:Clustering, Classification, and Retrieval Springer-Verlag New York, Inc. -2004.-244 p.

[Donchenko , Kirichenko,Serbaev, 2004] Donchenko V.S. Kirichenko M.F., Serbaev D.P. Recursive regression transformation and dynamical systems// Proceedings: of the Seventh International Conference “Computer Data analysis and Modeling: robustness and computer intensive methods”. – V.1. – September 6-10, 2004. – Minsk. – P. 147-151

[Donchenko, Kirichenko , Krivonos, 2007] [Donchenko V. Kirichenko M., Krivonos Yu. Generalizing of neural nets: Functional nets of special type// International Journal Information Theories and Applications. – N 3. –V.14. – 2007. – P. 259-266.

[Кириченко, Донченко,2007] Кириченко Н.Ф., Донченко В.С. Псевдообращение в задачах кластеризации.- Кибернетика и системный анализ.- 2007.- № 4.- с. 73 –92

[Friedman, Kandel,2000] Menahem Friedman Abraham Kandel Introduction to pattern recognition statistical, structural, neural and Fuzzy logic approaches world scientific publishing co. Pte. Ltd..-reprinted 2000.-329 p.

[Forster , Hölzl, 2004] John Forster and Verner Hölzl Applied Evolutionary economics and complex systems.-2004-- 293 p.

[Haykin, 2001] Haykin S. Neural Networks, A Comprehensive Foundation.-Prentice Hall, Upper Saddle River, New Jersey 07458. – 1999.– 842 p

[Kohonen,2001] Kohonen T. Self-Organizing Maps -3-d ed. - Tokyo: Springer, 2001.-501 p

[Kirichenko, Donchenko , Serbaev,2005] Kirichenko N.F., Donchenko V.S. Serbaev D.P. Nonlinear recursive nonlinear Transfomations: Dynamic sestems and Optimizations.//Cybernetics and Systen Ansalysis.– V.41, №3.– May 2005.– p.364-373.

[Kirichenko, 1997] N.F. Kirichenko. Analytical Representation of Perturbation of Pseudoinverse Matrices. Cybernetics and Systems Analysis. Vol.33, Number 2. P.230-239. March-April 1997.

[Nashed,1976] Nashed M. Zuhair ,Votruba G.F., editors. A Unified Operator Theory of Generalized Inverse.//Proceedings of an Advanced Seminar Sponsored by the Mathematical Research Center, The University of Wisconsin, Madison, October 8-10, 1973. – New York, Academic Press, 1976.–1200 p.

[Vapnik, 1998] Vapnik, V.N. Statistical Learning Theory. New York: Wiley, 1998

[Yeates, Wakefield, 2004] Donald Yeates, Tony Wakefield Systems Analysis And Design . -

Pearson Education. – 2004. - 499 p.

[Ivachnenko, 1969] Івахненко А.Г. Системи розпізнавання і автоматичного управління, що Самоорганізующиеся. – К.: Техніка, 1969.–395 с.

[Бублик, Кириченко,1975] Бублик Б.Н., Кириченко Н.Ф. Основы теории управления. – К.: Высшая школа, 1975.–328 с.

[Донченко,2011] Владимир Донченко Евклидовы пространства числовых векторов и матриц: конструктивные методы описания базовых структур и их использование.// International Journal “Information technologies & Knowledge”.- 2011.- Vol. 5.- Number 3.-P.203-216.

[Кириченко,1997] Н.Ф. Кириченко Н.Ф. Аналитическое представление возмущений псевдообратных матриц.// Кибернетика и системный анализ. – 1997. – № 2. – С. 98-107.

[Кириченко, Лепеха, 2002] Кириченко Н.Ф., Лепеха Н.П. Применение псевдообратных и проекционных матриц к исследованию задач управления, наблюдения и идентификации.// Кибернетика и системный анализ. – 2002. – № 4. – С. 107-124.

[Кириченко, Крак, Полищук, 2004] Кириченко Н. Ф., Крак Ю. В., Полищук А.А. Псевдообратные и проекционные матрицы в задачах синтеза функциональных преобразователей.// Кибернетика и системный анализ –2004.–№3.

[Кириченко , Донченко, Сербаев, 2005] Кириченко Н.Ф., Донченко В.С., Сербаев Д.П. Нелинейные рекурсивные регрессионные преобразователи: динамические системы и оптимизация.// Кибернетика и системный анализ.– №3, 2005.– С. 58-68.

[Кириченко, Кривонос, Лепеха 2007] Кириченко Н.Ф., Кривонос Ю.Г., Лепеха Н.П. Синтез систем нейрофункциональных преобразователей в решении задач классификации. // Кибернетика и системный анализ. – 2007. – №3. С. 47-57

[Кириченко, Донченко, Кривонос, Крак, Куляс, 2009] Кириченко М.Ф., Донченко В.С Кривонос Ю.Г., , Крак Ю.В.,., Куляс А.І. Аналіз та синтез ситуацій в системах прийняття рішень.- Київ: Наукова думка.-2009. 336с.

[Кириченко, Донченко,2005] Кириченко М.Ф., Донченко В.С. Задача термінального спостереження динамічної системи: множинність розв’язків та оптимізація//Журнал обчислювальної та прикладної математики. – 2005. –№5– С.63-78.

Authors' Information

Volodymyr Donchenko – *Professor, National Taras Shevchenko University of Kyiv. Volodymyrs’ka street, Kyiv, 03680, Ukraine; e-mail: voldon@bigmir.net.*

Yuri Krak– *Professor, National Taras Shevchenko University of Kyiv.*

Yuri Krivinos - *Deputy Director, Institute of cybernetics of NASU; Professor; academician, NASU*