2Recursive Regression Transformations and Dynamical Systems

V. Donchenko, M. Kyrychenko, D. Serbajev

Institute of Cybernetics, Kyiv National University - Faculty of Cybernetics
vsdon@unicyb.kiev.ua

Abstract

The report deals with the approximating problem for empirical dependencies. Special classes of transformations based on superposition and LSM are proposed. The transformations have special recursive structure and represented by discrete dynamic system.

1. Introduction

Discrete Dynamic systems [1] are described by the relations (1)-(2):

$$x(k+1)=f(x(k),u(k),k),\ k=0,...,N-1, \quad (1)$$

$$x(0)=x_{(0)}, \quad (2)$$

with Cost function

$$I(U)=\Phi(x(N))+\sum_{j=0}^{N-1}G(x(j),u(j),j), \quad (3)$$

U=(u(0),…,u(n)), (4)

f, Φ,G – scalar functions of vector arguments, as in classical: $x(k)\in R^m$, $k=\overline{0,N}$, $u(k)\in R^n$, $k=\overline{0,N-1}$, as well in generalized variant: $x(k)\in R^{m_k}$, $k=\overline{0,N}$, $u(k)\in R^{n_k}$, $k=\overline{0,N-1}$ and – beam dynamics also. In the last case the multiversion values for initial data are considered [2]. These systems are virtually the examples of the compound recurrent transformations with special parameters called a control. This peculiatrity of the systems especially rises above in the training of artificial neural network. The relation between dynamic systems and neural networks is established by the representation of the next theorem.

2. Neural nets and dynamics

Theorem 1. Neural network with N layers may be represented by the descrete dynamic system of the next type:

$$x(k+1)=\begin{pmatrix} F_1^{(k+1)}(w(k+1)^T x(k)) \\ ... \\ F_{l_{k+1}}^{(k+1)}(w(k+1)^T x(k)) \end{pmatrix},\ k=0,\dots,N-1,\text{де} \quad (5)$$

$F_1^{(k+1)}(z),...,F_{l_k}^{(k+1)}(z)$ – activation functions for the neurons of the layer with the number k, l_k – number of neurons of the correspondent layer, w(k), $k=\overline{1,N}$ – weights, $x(k)\in R^{l_k}$ – input, $x(k+1)\in R^{l_{k+1}}$ – output of the (k+1)-th layer, $k=\overline{0,N-1}$; x(0)– input, x(N) – output for neuronet on the whole.

Cost function J(w(1),…,w(N)) is determined by the relation:

$$J(w(1),\dots,w(N))=\sum_{i=1}^{M}\| y^{(i)}-x^{(i)}(N)\|^2,$$

where $y^{(i)}$ desired output value $x^{(i)}(N)$– real value of the net output for $x^{(i)}$ as input of the net, $i=\overline{1,M}$

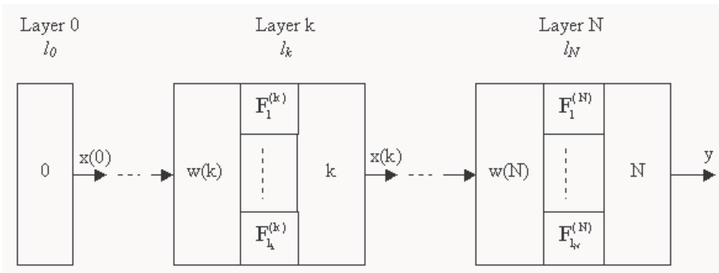


Fig 1.

Using of the neuronet for reconstructing the empirical dependance, represented by the learning sample stimulates of considerating the approximation problem in the classical or regression variant in its representation by the functions from certain class.

Example of the neuronet forces us to look at the function superposition as the main feature of correspondent functional transformation in the approximatig the dependance representing by the learning sample. The perspectives of such approach to using superposition are encouraging if taking into account the classical results due to A. Kolmogorov [3] and V.Arnold [4] relating superposition pole in representing function of many variables. We'd like to recollect also the results due to A. Ivachnenko [5] relating so called MGUA.

Thus it is really actually to consruct the approximation method vor empirical functions based on superpositions. It is advicable also to preserve all advatages of Least Square Method.

3. Recursive schemes for the constructing the functional transformations of special type – RFT transform

It is proposed to choice the model of input-output relation within the recursive using of the standard element – so called Elementary Recursive Regressive Transformation (ERRT). The transformation on a whole we will denote as RFT_N – Recursive Functional Transformation with N ERRT.

3.1. ERRT

Definition. We will call by ERRT, which approximate the dependence prescribed by the learning sample $(x_1^{(0)}, y_1^0),...,(x_M^{(0)}, y_M^0)$, $x_i^{(0)} \in R^n$, $y_i^{(0)} \in R^m$, the nonlinear transformation $\Phi : R^n \to R^m$:
$Y=\Phi(x)=A_+(\Psi(Cx))$, (6)
where:

- C–(n×n)-matrix;
- Ψ – component-wise nonlianear map from R^n in R^n with scalar transformations from the finite set $\Im$, including identical one;
- A_+– trace-minimal solution of the matrix equation $AX_{C\Psi}$ =Y,

where $X_{C\Psi}$ is assembled from the columns $\Psi(C x_i^{(0)})$, Y – from the $y_i^{(0)}, i=\overline{1,M}$, which correspond to the elemennts of the learning sample.
ERRT vertually is the empirical regression for linear regression y on $\Psi(Cx)$, constructed by LSM.

3.2. Primary types of the connections within RFT

There are 3 types of the connections, represented by figures 2-4. We will distinguish them by the means of parameter CON: “parinput”, “paroutput” and “seq”.

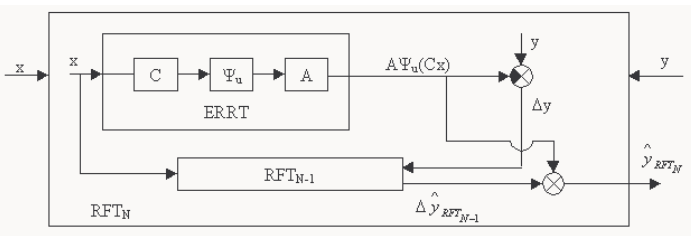


Fig 2. Output-parallel RFT- system with CON=parinput

3.3. Mathematical model of RFT-system – representation by the dynamic system

The RFT_N mathematical model may be represented by the dynamic system in which phase vector $\bar{x}(k)$, $k=\overline{0,N}$ describes RFT-subsystem with k ERRT-elemens, constructed by k-multiple recursive application ERRT-element: by superpositions in its. Vector $\bar{x}(k), k=\overline{0,N}$ consists of two parts, designated by x(k) and $x_\Sigma(k)$, $k=\overline{0,N}$:

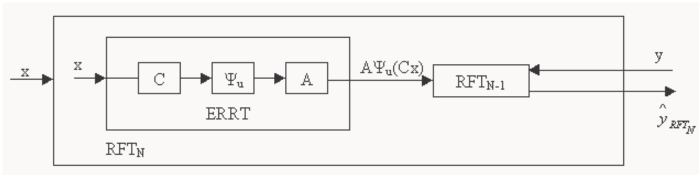


Fig 3. Output-parallel RFT-system with CON=paroutput

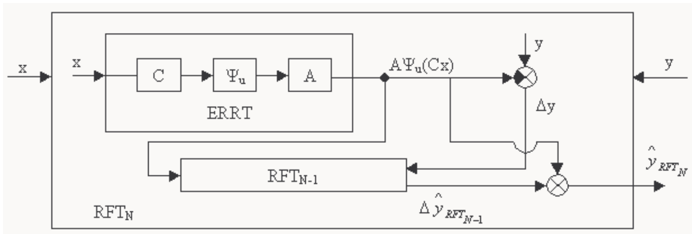


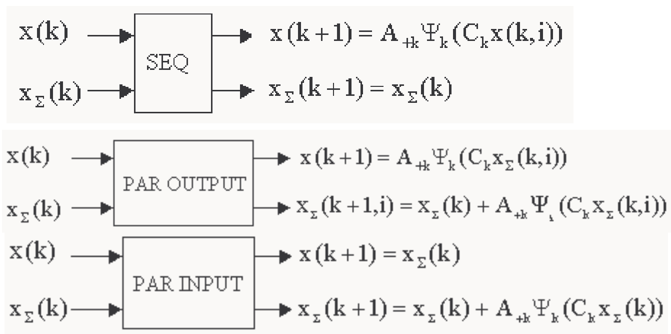
Fig 4. Output sequntial RFT-system: CON=seq

$$\bar{x}(k)=\begin{pmatrix} x(k) \\ x_\Sigma(k) \end{pmatrix}, k=\overline{0,N}.$$

First part x(k) – is the RFT_{k-1}-input for the system, substituted in ERRT, second $-X_\Sigma(k)$ – output for the RFT_k on whole, $k=\overline{0,N}$.

The recurrent relations between elements of the phase trajectory depens on the type of the connection: on the means of the parameter CON:

- Parallel with CON=parinput



- Parallel with CON=paroutput
- Sequential: CON=seq

Initial conditions are defined by the relations: $x(0)=x$, $x_\Sigma(0)=0$.

Thus the next statement is valid.

Theorem 2. RFT_N-system is equivalent by input-output relation to dynamics

$$\begin{pmatrix} x(k+1,i) \\ x_{\Sigma}(k+1,i) \end{pmatrix} = f\left(\begin{pmatrix} x(k,i) \\ x_{\Sigma}(k,i) \end{pmatrix}, A_{+k}, C_k, CON(k), k\right),$$

$$k = \overline{0, N-1}, \quad (6)$$

$$x(0) = x, \quad x_{\Sigma}(0) = 0. \quad (7)$$

Notice. Evidently, $A_{+k}, C_k, k = \overline{0, N-1}$ in (6)-(7) are the controll.

Theorem 3. Optimal choice problem for A_{+k}, C_k, $k = \overline{0, N-1}$ according to Least Square deviations on the leaning or test sample $(x_1^{(0)}, y_1^0), ..., (x_M^{(0)}, y_M^{(0)})$, $x_i^{(0)} \in R^n$, $y_i^{(0)} \in R^m$ for RFT_N-system with fixed structure is equivalent to optimal control problem for correspondent (6)-(7) beam dynamics with initial data $x_1^{(0)}, , ..., x_M^{(0)}, x_i^{(0)} \in R^n$. The controls U(k), $k = \overline{0, N-1}$ in the systems are the matrix pares: U(k)=(A_{+k}, C_k). $k = \overline{0, N-1}$, and the cost function is the sum of the square deviations RFT_N-outputs from the output elements of the learning sample.

Theorem 4. The derivatives of the cost function for the RFT approximation by the A_{+k}, C_k, $k = \overline{0, N-1}$ may be calculated as the correspondent derivatives from Hamilton function for dynamics (6)-(7) with correspondent cost function.

The solution of the problem of choosing Ψ_k, $k = \overline{1, N}$ one can find in [6].

4. References

[1] Бублик Б.Н., Кириченко Н.Ф. Основы теории управления.–Киев: Вища школа.1975.–328 с.

[2] Бублик Б.Н., Гаращенко Ф.Г., Кириченко Н.Ф. Структурно-параметрическая оптимизация и устойчивость динамических пучков.–Киев: Наукова думка.1985.–304 с.

[3] Колмогоров А.Н. О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных // Докл. АН СССР. – 1956. – Т. 108. – № 2. – С. 179-182.

[4] Арнольд В.И. О функциях трех переменных // Докл. АН СССР. – 1957. – Т. 114. – № 4. – С. 953-956.

[5] Ивахненко А.Г. Самоорганизующиеся системы распознавания и автоматического управления. – К.: Техника. – 1969. – 395 с.

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