GENERALIZING OF NEURAL NETS: FUNCTIONAL NETS OF SPECIAL TYPE

Volodymyr Donchenko, Mykola Kirichenko, Yuriy Krivonos

Abstract: Special generalizing for the artificial neural nets: so called RFT – FN – is under discussion in the report. Such refinement touch upon the constituent elements for the conception of artificial neural network, namely, the choice of main primary functional elements in the net, the way to connect them(topology) and the structure of the net as a whole. As to the last, the structure of the functional net proposed is determined dynamically just in the constructing the net by itself by the special recurrent procedure. The number of newly joining primary functional elements, the topology of its connecting and tuning of the primary elements is the content of the each recurrent step. The procedure is terminated under fulfilling "natural" criteria relating residuals for example. The functional proposed can be used in solving the approximation problem for the functions, represented by its observations, for classifying and clustering, pattern recognition, etc. Recurrent procedure provide for the versatile optimizing possibilities: as on the each step of the procedure and wholly: by the choice of the newly joining elements, topology, by the affine transformations if input and intermediate coordinate as well as by its nonlinear coordinate wise transformations. All considerations are essentially based, constructively and evidently represented by the means of the Generalized Inverse.

Keywords: Artificial neural network, approximating problem, beam dynamics with delay, optimization.

ACM Classification keywords:F.1.1.Models of Computation: Self modifying machines(neural networks; G.1.6. Optimization; H.1.m. Models and principles; I.2.6. Artificial Intelligence: learning, connectionism and neural nets.

Introduction

Artificial neurons nets are the technological elements, used in various applications ([Amit, 2002], [Veelenturf, 1995] for example) especially under model uncertainty. It may be approximation problem for the functions represented by its observations, the task of the control, classification problem and so on.

The power of the artificial neural nets (ArtNN) substantially specified by the using virtually the composition of the functions [Donchenko, Kirichenko, Serbaev, 2004]. The might of this instrument was certified by [Kolmogorov, 1966] and [Arnold, 1967].

But some faults of the ArtNN are obvious, namely, the constraints on the primary functional elements represented by neurons, constraints on topology, constraints on the optimizations, properly in the appearance of the Back Propagation.

In the report the attempt is undertaking to extend the ArtNN up to Functional nets taking as the primary the special functional element – so called ERRT: elementary recursive regressive transformation, spreading the topology and introducing "natural" optimizing parameter.

The fulcrum for the implementing this attempt arises from the theory of Generalized Inverse Matrixes (for ex. [Алберт, 1976]) developed by one of the authors and represented in [Кириченко, Лепеха, 2002].

The approach proposed is realized in the conception of so called Recursive nonlinear Functional Transformation - Functional Net (RFT-FN. The idea of such type transformations in the variant of inverse recursion was proposed in [Кириченко, Крак, Полищук, 2004]], focused on the optimizing procedure for ERRT. We represent here another variant of RFT - FN – the variant of so called inverse recursion, introduced and discussed in the paper [Donchenko, Kirichenko, 2005] by the authors. As it has been already noticed earlier, RFT – FN embodies the main ideas of the classical ArtNN: using the composition of standard primary functional elements (artificial neurons – a.n.) connected according some topology for constructing a complex object. The primary elements of the composition: a.n. – represents mathematically comparatively simple function: a composition of linear function – linear functional for input multi-dimensional variable, – and simple standard scalar function. Conception of the RFT-FN leaves the main idea of ArtNN unchangeable: that the idea of constructing a complex objects through a composition of standard simple ones. But it is proposed and implemented some substantial refinements.

The main of them are actually in the next:

- Expanding the domain of the feasible functions for the basis standard elements and introducing the special procedure for the adjusting for the newly connecting ERRT;

Introducing the special recursive dynamic procedure for the constructing the RFT-FN. The term "dynamic" means the nor a priory structure is fixed for the RFT-FN, but it determines exclusively by the quality of the constructions, characterized by some "natural" functional of the quality, the residuals in the approximation for example. So the termination of the recursive procedure of the constructing the RFT-FN determines dynamically in the in the course of the procedure by itself.

- Expanding the variants of the feasible connections for the newly connecting elements and its number.

There some more additional enhancements for the ArtNN, proposed and implemented in the conception of the RFT – FN. These are: coordinate wise nonlinear transformations of the input and intermediate inner coordinate along with the linear transformations of them.

All the refinements make it possible to optimize the constructing of the RFT-FN on the each recursive step and wholly by the next ways:

- By the choice of the feasible functions for the ERRT, by the choice of linear transformations of its coordinates and coordinate wise nonlinear transformations of the coordinate for the input variables for the newly connected ERRT: single or a number of them;

- By the choice of a topology for newly connected elements. There are three mains types of the connections for the newly connecting elements: parallel by input (parinput), parallel by output (paroutput) or sequential (seq).

The implementation of the RFT–FN conception will be demonstrated on the approximation problem for the functions, represented by its observations. Draw the attention of the reader that this problem is represented widely in the publications as in deterministic as well in statistical enunciations. As regarding the last, we refer to [Линник, 1962], [Вапник, 1979], [Ивахненко, 1969].

1. General Conception of the Functional Net: RFT -FN

The RFT-FN constructing procedure conclude in joining the recurrently EERT to RFT-FN already has been constructed during previous steps in compliance with one of the three certain types of connecting,

These types will be denoted as parinput, paroutput and seq. The connecting types correspond to natural transformations of input signal: parallel by input or output, and also sequential.

1.1. Description of the Primary Functional Elements: ERRT

The basic constructive element for RFT-FN is the ERRP-element [9], which is defined as a map from R^{n-1} in R^m designed according to the next form:

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

It approximates the dependence, represented by the learning 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 is (n×n))-matrix, defining an affine map between R^{n-1} and R^m, fixed when synthesizing the ERRT;

Ψ_u – coordinate-wise nonlinear map from R^n in R^n; each of nonlinear real functions u_i, $i = \overline{1, n}$, transforming the coordinates, belongs to finite set $\Im$ of the functions. This set is fixed, but open for extension. The set $\Im$ include also identity function. We will consider each of such functions to be smoothly enough; one chose the nonlinear map Ψ to minimize the discrepancy between input and output on the learning sample;

$A+$ – trace-norm minimal solution of the next matrix equation

$$AX_{\Psi_u C} = Y, \qquad\qquad (2)$$

in which $X_{\Psi_u C}$ – matrix assembled from vectors-columns $\Psi_u(C\begin{pmatrix} x_i^{(0)} \\ 1 \end{pmatrix}) = \Psi_u(z_i^{(0)})$ and matrix Y –

from $y_i^{(0)}, i = \overline{1, M}$

1.2. Recursive Procedure and Topology of Connections in RFT – FN Constructing

Composition-recursion in function-building procedure in the proposed below variant of direct recursion will be considered in generalized version. In such version the number k_m of newly joined ERRT may be more, than one.

Total number of ERRP used will be denoted by T: $T = \sum_{m=1}^{N} k_m$, N being the number of recursive calls of the procedure.
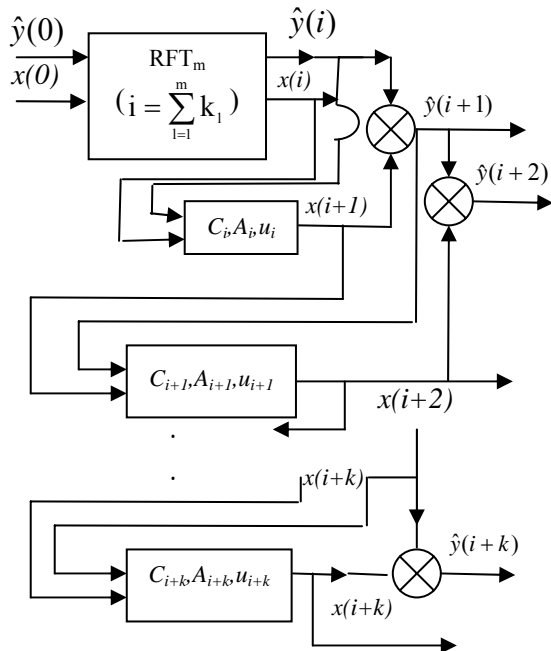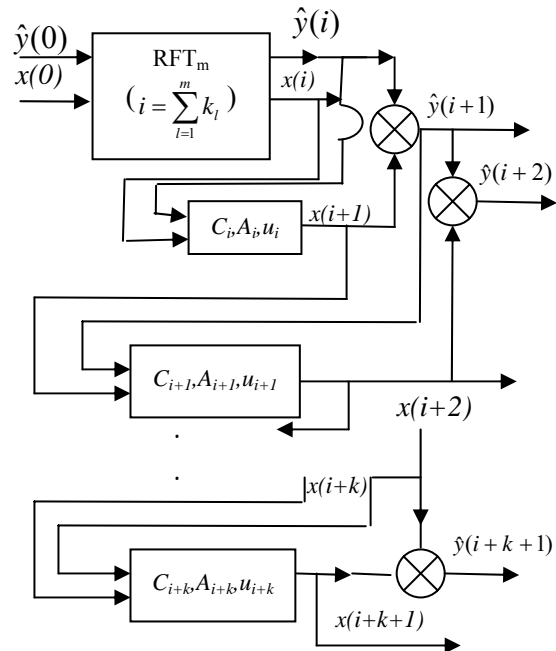

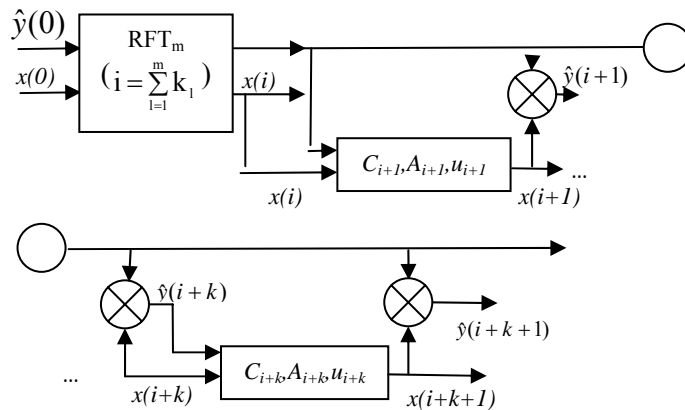Fig. 1. Parinput

Fig. 2. Paroutput

Fig. 3. Seq

Direct recursion is represented by the figures 1, 2, 3 and corresponding equations depending on type of the joining.

- Parinput (fig. 1)

The input-output equations represented parinput topology in the recursive procedure are tot the next form:

$$x(i+j) = A_{+i+j-1} \Psi_{u_{i+j-1}} (C_{i+j-1} x(i)),$$

$$\hat{y}(i+j) = \hat{y}(i+j-1) + A_{+i+j-1} \Psi_{u_{i+j-1}} (C_{i+j-1} \cdot x(i)),$$

$$i = \sum_{l=1}^{m} k_l, \ j = \overline{1, k_{m+1}}$$

(3)

- Paroutput (fig. 2)

Paroutput topology is calculated by the next input-output equations:

$$x(i+j) = A_{+i+j-1} \Psi_{u_{i+j-1}} (C_{i+j-1} x(i+j-1)),$$

$$\hat{y}(i+j) = \hat{y}(i+j-1) + A_{+i+j-1} \Psi_{u_{i+j-1}} (C_{i+j-1} \cdot x(i+j-1)),$$

$$i = \sum_{l=1}^{m} k_l, \ j = \overline{1, k_{m+1}} \ .$$

(4)

- Seq (fig.3)

And at last for sequential type we have correspondingly:

$$x(i+j) = A_{+i+j-1} \Psi_{u_{i+j-1}} (C_{i+j-1} x(i+j-1)),$$

$$\hat{y}(i+j) = \hat{y}(i) + A_{+i+j-1} \Psi_{u_{i+j-1}} (C_{i+j-1} \cdot x(i+j-1)),$$

$$i = \sum_{l=1}^{m} k_l, \ j = \overline{1, k_{m+1}}$$

(5)

2. Special Class of Beam Dynamics with Delay

The optimization for RFT – FN as it is follows from (3)-(5)is reduced virtually to solving the optimization problem for the beam dynamics of special type, determined below. Namely, we will introduce and consider two classes of special discrete dynamic systems with delay named below simple and combined. Classical results about conjugate systems and Hamilton functions will be extended on the systems introduced as well as the results about functional differentiating respectively controls.

2.1. Special Class of Beam Dynamics with Delay : Basic Definitions

These two types of beam dynamics: with simple delay and combined – are defined in the next way:
- simple delay:

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

(6)

combined:

$$x(j+1) = f(x(j), x(j-s(j)), u(j), j) \ ; \quad j = \overline{0, N-1} \ ;$$

(7)

set of the initial states Ini$= \left\{ x_1^{(0)}, ..., x_M^{(0)} \right\}$: $x(0) = x^{(0)} \in$ Ini;

- functional on the set of the trajectories

$$I(U) = \sum_{x^{(0)} \in Ini} \Phi(x(N)),$$ (8)

delay function $s(j) \in \{2,..., j\}$, $s(0)=0$, $s(j) \in \{2,..., j\}$, $j = \overline{0, N-1}$.

Simple systems are defined by the collection of the functions $f(z, u, j)$, $j = \overline{0, N-1}$ and combined ones – by the $f(z, v, u, j)$, $j = \overline{0, N-1}$.

2.2. Conjugate Systems and Hamilton Function

Given the system dynamics with delay: simple or combined – define the conjugate systems and the Hamilton functions depending on the type of the delay beam dynamics.

Simple delay:

- Conjugate system $p(k)$, $k = \overline{N,0}$

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

$$p(k) = \sum_{j \in J(k)} grad_{x(k)}\{p^T(j+1)f(x(j),u(j),j)\},$$

$$J(k) = \{j : j - s(j) = k, j \geq k\}, \quad k = \overline{N-1,0};$$

- Hamilton function:

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

Combined delay:

- Conjugate system $p(k)$, $k = \overline{N,0}$

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

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

$J(k)$ is the same as for simple systems, $k = \overline{N-1,0}$;

- Hamilton function:

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

2.3. Gradient in Beam Dynamics with Delay

The classical results take place for the beam dynamics with delay within the classical assumptions as to $f(z, u, j)$ or $f(z, v, u, j)$, $j = \overline{0, N-1}$ and Φ. These results are captured in the next two theorems.

Theorem 1. For the simple delay beam dynamics gradients respectively controls are represented by the next relations

$$grad_{u(k)}I(U) = -grad_{u(k)}\sum_{i=1}^{M}H^{(i)}(p^{(i)}(k+1)x^{(i)}(k), u(k), k), \quad k = \overline{0, N-1}.$$

Theorem 2. For the combined delay beam dynamics gradients respectively controls are represented by the next relations

$$grad_{u(k)}I(U) = -\sum_{i=1}^{M} grad_{u(k)} H^{(i)}(p^{(i)}(k+1), x^{(i)}(k), x^{(i)}(k-s(k), u(k), k)), \quad k = \overline{0, N-1}.$$

Index i: $i = \overline{1, M}$ corresponds to trajectories with initial states $x_i^{(0)} \in Ini$.

3. Functional Nets and Beam Dynamics with Delay

Combined delay beam dynamics are very important regarding their role in representation of RFT-FN constructions.

Theorem 3. RFT-FN – predictor with the direct N recursions in using $k_m, m = \overline{1, N}$ ERRT respectively can be represented by the combined delay beam dynamics on the time interval $\overline{0, T}$, $T = \sum_{l=1}^{N} k_l$. The elements of such representations are constructive, depending on the type of the joining. The quality functional of the system is of the next form

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

where $z_2(T)$ is one of two output components for the beam dynamics.

4. Functional Nets Optimal Design

Theorem 3 enables to choose optimally $C_0, C_1, ..., C_{T-1}$ for RFT – FN.

Theorem 4. Under assumption that any element from $\Im$ has the continuous second-order derivation, an RFT-FN is feasible to be constructively optimized by gradient methods respectively matrixes C.

Conclusion

Special kind of the "functional nets": so called RFT – FN, generalizing classical functional nets, namely, artificial neural nets, has been proposed and investigated in the report. The RFT – FN permit multilevel optimization. First level optimization is the optimization in the primary functional element and the pseudo inverse is principally in this stage. Another principal level of the optimization is the optimization due to optimization in the beam dynamics.

Bibliography

[Amit, 2002] Y. Amit., 2D Object Detection and Recognution: models, algorithms and networks.–The MIT Press, Cambridge, Massachusetts.– 2002.–306 p.

[Veelenturf,1995] L.P.J.Veelenturf, Analysis and applications of artificial Neural Networks.–Prentice Hall (UK).– 1995.–259 p.

[Donchenko, Kirichenko, Serbaev,2004] V.S. Donchenko, N.F.Kirichenko,D.P.Serbaev. Recursive regression transformations and dynamical systems.// Proceedings of the Seventh International Conference "Computer Data Analysis and modelling: robustness and computer intensive methods".– V.1.–September 6-10, 2004. – Minsk (Belarus).– p.147-151.

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

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

[Алберт,1977] А.Алберт Регрессия, псевдоинверсия, рекуррентное оценивание.– Пер.с англ.– М.: Наука.–1977.–305c.

[Кириченко,Лепеха,2002] Н.Ф. Кириченко, Н.П.Лепеха 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] Н.Ф.Кириченко, Ю.В. Крак.А.А. Полищук Псевдообратные и проекционные матрицы в задачах синтеза функциональных преобразователей.// Кибернетика и системный анализ –2004.–№3.

[Donchenko, Kirichenko,2005.] V.S Donchenko, N.F. Kirichenko Generalized Inverse in Control with the constraints.// Cybernetics and System Analysis.– v.39 (6) November-December– 2003.–p.854-861(USA).

[Линник,1962] Ю.В.Линник Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. Изд. 2-е. – М.:Физматгиз.– 1962.– 349 с.

[Вапник,1979] В.Н. Вапник. Восстановление зависимостей по эмпирическим данным. – М.: Наука. – 1979. – 447 с.

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

Authors' Information

Volodymyr Donchenko – Kyiv National Taras Shevchenko University (Ukraine), Professor, e-mail: voldon@unicyb.kiev.ua

Mykola Kirichenko – Institute of the Cybernetics, National Academy of Sciences (Ukraine), Professor

Yuriy Krivonos – Institute of the Cybernetics, National Academy of Sciences (Ukraine), Member Correspondent of the National Academy of Sciences.

NEURAL NETWORK BASED OPTIMAL CONTROL WITH CONSTRAINTS

Daniela Toshkova, Georgi Toshkov, Todorka Kovacheva

Abstract: *In the present paper the problems of the optimal control of systems when constraints are imposed on the control is considered. The optimality conditions are given in the form of Pontryagin's maximum principle. The obtained piecewise linear function is approximated by using feedforward neural network. A numerical example is given.*

Keywords: *optimal control, constraints, neural networks*

ACM Classification Keywords: *I.2.8 Problem Solving, Control Methods, and Search*

Introduction

The optimal control problem with constraints is usually solved by applying Pontryagin's maximum principle. As it is known the optimal control solution can be obtained computationally. Even in the cases when it is possible an analytical expression for optimal control function to be found, the form of this function is quite complex. Because of that reason the possibilities of using neural networks for solving the optimal control problem are studied in the present paper.

The ability of neural networks to approximate nonlinear function is central to their use in control. Therefore it can be effectively utilized to represent the regulator nonlinearity. Other advantages are their robustness, parallel architecture.

Lately, different approaches are proposed in the literature treating the problem of constrained optimal control for using neural networks. In [Ahmed 1998] a multilayered feedforward neural network is employed as a controller. The training of the neural network is realized on the basis of the so called concept of Block Partial Derivatives. In [Lewis 2002] a closed form solution of the optimal control problem with constraints is obtained solving the associate Hamilton-Jacobi-Bellman (HJB) equation. The solution of the value function of HJB equation is approximated by using neural networks.