Theory of Functional Systems, Adaptive Critics and Neural Networks

Vladimir G. Red'ko,
Institute of Optical Neural Technologies, Russian Academy of Science
Vavilova Str., 44/2, Moscow, 119333, Russia.
redko@iont.ru

Danil V. Prokhorov
Ford Research and Advanced Engineering, Ford Motor Company
2101 Village Rd., MD 2036, Dearborn, MI 48124, U.S.A.
dprokhor@ford.com

Mikhail S. Burtsev
M.V. Keldysh Institute for Applied Mathematics, Russian Academy of Science,
Miusskaya Sq., 4, Moscow, 125047, Russia
mbur@narod.ru

***Abstract* — We propose a general scheme of intelligent adaptive control system based on the Petr K. Anokhin's theory of functional systems. This scheme is aimed at controlling adaptive purposeful behavior of an animat (a simulated animal) that has several natural needs (e.g., energy replenishment, reproduction). The control system consists of a set of hierarchically linked functional systems and enables predictive and goal-directed behavior. Each functional system includes a neural network based adaptive critic design. We also discuss schemes of prognosis, decision making, action selection and learning that occur in the functional systems and in the whole control system of the animat.**

I. Introduction

In the early 1990s, the animat approach to artificial intelligence (AI) was proposed. This research methodology implies understanding intelligence through simulation of artificial animals (*animats*) in progressively more challenging environments [1, 2]. The main goal of this field of research is "designing animats, i.e., simulated animals or real robots whose rules of behavior are inspired by those of animals. The proximate goal of this approach is to discover architectures or working principles that allow an animal or a robot to exhibit an adaptive behavior and, thus, to survive or fulfill its mission even in a changing environment. The ultimate goal of this approach is to embed human intelligence within an evolutionary perspective and to seek how the highest cognitive abilities of man can be related to the simplest adaptive behaviors of animals" [3].

In this paper we propose a general scheme of an animat control system based on the theory of functional systems. This theory was proposed and developed in the period 1930-1970s by Russian neurophysiologist Petr K. Anokhin [4] and provides general schemes and regulatory principles of purposeful adaptive behavior in biological organisms. In addition to the theory of functional systems, we use the important approach to development of intelligent control systems called adaptive critic designs (ACD) [5-10]. Using neural network based ACD, we propose a hierarchical control system for purposeful adaptive behavior. This paper, along with [11], represents our first attempt to design an *animat brain* from the principles of the theory of functional systems. Some ideas of the current work are similar to those of the project "Animal" initiated by Bongard et al. in the 1970s [12], as well as to those of proposals by Werbos on brain-like multi-level designs [13]. Schemes and designs proposed here are intended to provide a framework for a wide range of computer simulation models.

Section II reviews the Anokhin's theory of functional systems. Section III describes an ACD used in the current work. Section IV contains the description of the whole control system. Section V concludes the paper.

II. Anokhin's Theory of Functional Systems

Functional system was proposed in 1930s as "a complex of neural elements and corresponding executive organs that are coupled in performing defined and specific functions of an organism. Examples of such functions include locomotion, swimming, swallowing, etc. Various anatomical systems may participate and cooperate in a functional system on the basis of their synchronous activation during performance of diverse functions of an organism" [14].

Functional systems were put forward by P.K. Anokhin as an alternative to the predominant concept of reflexes. Contrary to reflexes, the endpoints of functional systems are not actions themselves but adaptive results of these actions. This conceptual shift requires understanding of biological mechanism for matching results of actions to adaptive requirements of an organism, which are stored as anticipatory models in the nervous system. A biological feedback principle was introduced in the scheme of the functional system in 1935 as a *backward afferentation* flowing through different sensory channels to a central nervous system after each action [15]. An anticipatory neural template of a required result placed into memory before each adaptive action was called an *acceptor* of the result of action [16]. The term acceptor carries two meanings derived from its Greek root: (1) acceptor as a receiver of the action's feedback, and (2) acceptor as a neural template of the goal to be compared with feedback and, in the case of positive match between the model and feedback, followed by the action's acceptance.

In contrast to reflexes, which are based on linear spread of information from receptors to executive organs through the central nervous system, functional systems are self-organizing non-linear systems composed of synchronized distributed elements. The main experimental issues of research on functional systems amounted to understanding how this self-organization is achieved and how information about the goal, plans, actions and results is represented and processed in such systems. These studies led to creation of the conceptual scheme of stages of adaptive behavioral acts shown in Fig. 1.

The main stages of the functional system operation are (see Fig.1):

1) afferent synthesis,
2) decision making,
3) generation of the acceptor of the action result,
4) generation of the action program (efferent synthesis),
5) performance of an action,
6) attainment of the result,
7) backward afferentation (feedback) to the central nervous system about parameters of the result,
8) comparison of the result with its model generated in the acceptor of the action result.

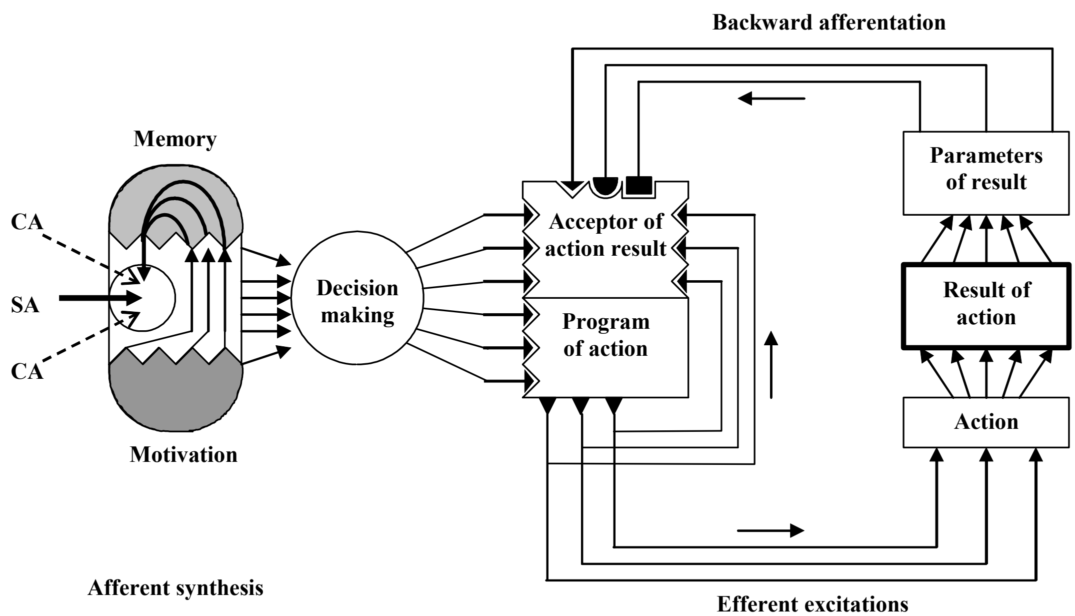

Fig. 1. General architecture of a functional system. SA is starting afferentation, CA is contextual afferentation. Operation of the functional system includes: 1) preparation for decision making (afferent synthesis), 2) decision making (selection of an action), 3) prognosis of the action result (generation of acceptor of action result), 4) backward afferentation (comparison between the result of action and the prognosis). See the text for details.

Operation of the functional system is described below.

The afferent synthesis precedes each behavioral action and involves integration of neural information from a) dominant motivation (e.g., hunger), b) environment (including contextual and conditioned stimuli), and c) memory (including evolved biological knowledge and individual experience). The afferent synthesis can occupy a substantial time and involve cycles of reverberation of signals among various neural elements.

The afferent synthesis ends with decision making, which means a reduction of redundant degrees of freedom and selection of a particular action in accordance with a dominant need, organism's experience and environmental context.

The efferent synthesis is preparation for the effectory action and further reduction of the excessive degrees of freedom by selection of actions most suitable for the current organism's position in space, its posture, information from proprioreceptors, etc.

The acceptor for the action result is being formed in parallel with the efferent synthesis. This produces an anticipatory model of the required result of action. Such a model includes formation of a distributed neural assembly that stores various (i.e., proprioreceptive, visual, auditory, olfactory) parameters of the expected result.

Performance of every action is accompanied by backward afferentation. If parameters of the actual result are different from the predicted parameters stored in the acceptor of action result, a new afferent synthesis is initiated. In this case, all operations of the functional system are repeated until the final desired result is achieved.

Thus, operation of the functional system has a cyclic (due to backward afferent links), self-regulatory organization.

A separate important branch of the general functional system theory is the theory of systemogenesis that studies mechanisms of functional systems formation during 1) evolution, 2) individual or ontogenetic development, and 3) learning. In the current paper we consider learning only, i.e., improvement of functional systems operation based on individual experience.

It should be stressed that the theory of functional systems was proposed and developed in order to interpret a volume of neurophysiological data. This theory was formulated in very general and intuitive terms. We are only in the beginning of formalization of the theory of functional systems by means of mathematical and computer models [11,17]. Though formal powerful models of this theory are not created yet, it is supported by numerous experimental data and provides an important conceptual basis to understanding brain operation. This theory could help us to understand neurophysiological aspects of prognosis, prediction, and creation of causal relationship among situations in animal minds. The theory of functional systems could also serve as a conceptual foundation for modeling of intelligent adaptive behavior.

We propose a simple formalization of the functional system, and we use this formalization for designing the whole animat control system. Our functional system includes the following important features of its biological prototype: a) prognosis of the action result, b) comparison of the prognosis and the result, and c) correction of prognosis mechanism via learning in appropriate neural networks. Our functional system utilizes one of the possible schemes of ACD, as described in the next section.

III. Neural Network Based Adaptive Critic Design

Our adaptive critic scheme consists of two neural network based blocks: model and critic (see Fig.2). For simplicity, we assume that both neural networks are differentiable feed-forward multilayer perceptrons, and that their derivatives can be computed via the well known back-propagation algorithm. Depending on the problem, other network architectures with their associated training methods may be more suitable to employ. In particular, recurrent neural networks can be used instead of feed-forward perceptrons in order to ensure short-term memory in the form of neural exitation activity.

We suppose that our adaptive critic serves to select one from several actions. For example, for movement control the actions can be move forward, turn left, turn right. The animat in any moment t should select one of these actions.

The goal of our adaptive critic is to maximize stochastically utility function $U(t)$:

$$U(t) = \sum_{j=0}^{\infty} \gamma^{j} r(t_{j}), \quad t = t_0, t_1, t_2, \ldots, \qquad (1)$$

where $r(t_j)$ is a particular reinforcement (reward, $r(t_j) > 0$, or punishment, $r(t_j) < 0$) obtained by the adaptive critic at the moment t_j, and γ is the discount factor ($0 < \gamma < 1$). In general, the difference $t_{j+1} - t_j$ may be time varying, but for notational simplicity we assume that $\tau = t_{j+1} - t_j = \text{const}$.

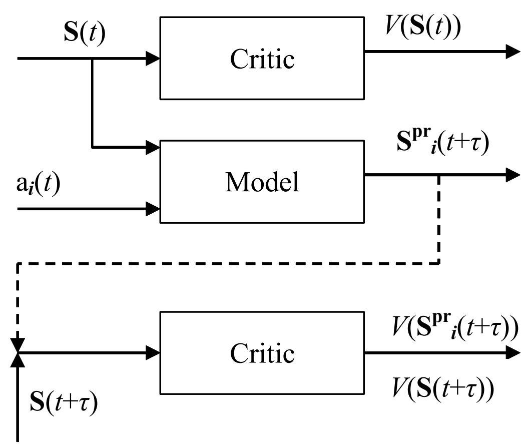

Fig. 2. The adaptive critic scheme used in our functional system. The model predicts the next state $\mathbf{S}^{\mathbf{pr}}_i(t+\tau)$ for all possible actions a_i, i=1,2,…, n_a. The current state $\mathbf{S}(t)$, the prediction $\mathbf{S}^{\mathbf{pr}}_i(t+\tau)$ and the next state $\mathbf{S}(t+\tau)$ are fed into the critic (the same neural network is shown in two consecutive time moments) to form the corresponding values $V(\mathbf{S}(t))$, $V(\mathbf{S}^{\mathbf{pr}}_i(t+\tau))$ and $V(\mathbf{S}(t+\tau))$ of the state value function.

The model has two kinds of inputs: 1) a set of inputs characterizing the current state $\mathbf{S}(t)$ (signals from external and internal environments of animat), and 2) a set of inputs characterizing actions. We assume that any possible action a_i is characterized by its own combination of inputs and that the number of possible actions is small. The role of the model is to make predictions of the next state for all possible actions a_i, i=1,2,…, n_a.

The critic is intended to estimate the state value function $V(\mathbf{S})$ for the current state $\mathbf{S}(t)$, the next state $\mathbf{S}(t+\tau)$ and its predictions $\mathbf{S}^{\mathbf{pr}}_i(t+\tau)$ for all possible actions.

At any moment of time the following operations are performed:

1) The model predicts the next state $\mathbf{S}^{\mathbf{pr}}_i(t+\tau)$ for all possible actions a_i , $i=1,\ldots, n_a$.

2) The critic estimates state value function for both the current state $V(t) = V(\mathbf{S}(t))$, and predicted states $V^{\mathbf{pr}}_i(t+\tau) = V(\mathbf{S}^{\mathbf{pr}}_i(t+\tau))$. The values V are estimates of the utility function U.

3) The ε-greedy rule is applied [18], and the action is selected as one of the two alternatives below:

$k = \arg\ \max_i\{V(\mathbf{S}^{\mathbf{pr}}_i(t+\tau))\}$ with probability 1- ε,

OR

k is index of an arbitrary state-meaningful action randomly chosen with probability ε,

where k is index of selected action a_k .

4) The action a_k is carried out.

5) The current reinforcement $r(t)$ is received (or estimated) before or after the transition to the next time moment $t+\tau$ occurs. The next state $\mathbf{S}(t+\tau)$ is observed and is compared with the predicted state $\mathbf{S}^{\mathbf{pr}}_k(t+\tau)$. The neural network weights $\mathbf{W}_M$ of the model may be adjusted to minimize the prediction error:

$$\Delta\ \mathbf{W}_M = \alpha_M\ \mathrm{grad}_{\mathbf{W}M}(\mathbf{S}^{\mathbf{pr}}_k(t+\tau))^{\mathrm{T}}(\mathbf{S}(t+\tau) - \mathbf{S}^{\mathbf{pr}}_k(t+\tau))\ , (2)$$

where α_M is the learning rate of the model.

6) The critic estimates $V(\mathbf{S}(t+\tau))$. The temporal-difference error is calculated:

$$\delta(t) = r(t) + \gamma\ V\,(\mathbf{S}(t+\tau)) - V\,(\mathbf{S}(t))\ . \qquad (3)$$

7) The weights $\mathbf{W}_C$ of the critic neural network may be adjusted:

$$\Delta\ \mathbf{W}_C = \alpha_C\ \delta(t)\ \mathrm{grad}_{\mathbf{W}C}(V(t))\ , \qquad (4)$$

where α_C is the learning rate of the critic. The gradients $\mathrm{grad}_{\mathbf{W}M}(\mathbf{S}^{\mathbf{pr}}_k(t+\tau))$ and $\mathrm{grad}_{\mathbf{W}C}(V(t))$ mean derivatives of the outputs of the networks with respect to all appropriate weights. Instead of (4), we can use a more complex temporal-difference learning scheme from [18].

The described adaptive critic design is the core of our functional system. Many functional systems form the entire animat control system, as described in the next section.

IV. Design of Animat Control System

We suppose that animat control system has a hierarchical architecture (Fig. 3). The basic element of this control system is our adaptive critic based functional system (FS). Each FS makes a prognosis of the next state and selects actions in accordance with the currently dominant animat need and the current state.

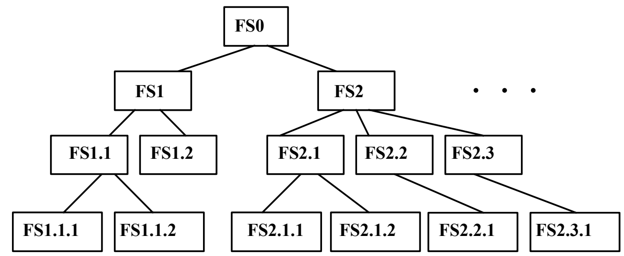

Fig. 3. Hierarchical structure of the animat control system. FS* represents a separate functional system.

The highest level of the hierarchy (FS0) corresponds to survival of the simulated animal. The next level (FS1, FS2,…) corresponds to the main animal needs (e.g., energy replenishment, reproduction, security, knowledge acquisition). Lower levels correspond to tactical goals and sub-goals of behavior.

Control commands can be delivered from high levels (super-system levels) to low levels (sub-system levels) and returned back. They include activation commands (delivered from a super-system to its sub-system) and report commands (delivered from a sub-system to a super-system). These commands enable propagation of activity through the entire control system. For simplicity in this work we suppose that, at any time moment, only one FS is active.

The detailed structure of our FS is shown in Fig. 4. In its core is the ACD described in Section III. The main differences between operation of the ACD and that of the FS are 1) the FS additionally forms commands for sub-systems and reports to super-systems, and 2) comparison between the prognosis $\mathbf{S}^{\mathbf{pr}}_k(t+\tau)$ and the result $\mathbf{S}(t+\tau)$ can be postponed until the moment $t+\tau$, when reports from sub-systems are received (see below for details). Links of the given FS to a super-system/sub-system are shown by vertical solid/dotted arrows.

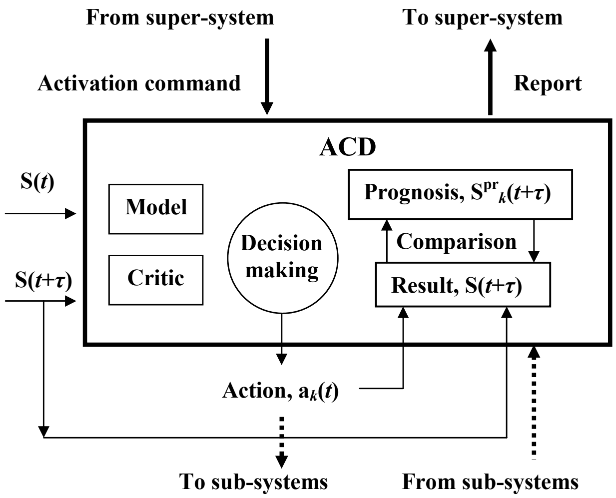

Fig. 4. Scheme of the functional system based on the adaptive critic design. The symbolic notation is the same as in Fig. 2 (see Section III).

We assume the following scheme of operation of the FS within the animat control system. The FS is activated by the command from a super-system. The model and the critic operate in the same manner as described in Section III, and the action a_k is selected. Further operation depends on the kind of action a_k. Some actions are commands for effectors, and these actions are executed immediately; in this case $\tau = \tau_{min}$, the smallest time interval allowed in the system. Then the animat reinforcement r is received from external or internal environment, and the ACD neural network learning is carried out.

Another type of actions is a command for sub-systems. For such an action, the command to activate a certain sub-system is delivered (which sub-system to activate is determined by the selected action a_k). In this case the comparison of prognosis and result, the estimation of reinforcement value r and learning of the neural networks are postponed until the moment $t + \tau$, where $\tau > \tau_{min}$.

Thus, operations in the FS are actually the same in both cases, the main difference being the moment of learning $t+\tau$ (e.g., τ can be much larger than τ_{min}). After accomplishing all these operations, the FS sends a report about completion of its activity to an appropriate super-system.

The described mode of FS operation represents the ordinary mode of functioning. We also specify the extraordinary mode taking place if the prognosis differs substantially from the actual result: $\| \mathbf{S}^{\mathbf{pr}}_k(t_j) - \mathbf{S}(t_j) \| > \Delta > 0$, where $\| . \|$ denotes some norm, e.g. Euclidean. We suppose that in the extraordinary mode the values ε (the probability of choosing a random (but meaningful) action; see Section III) of the given FS and its sub-systems increase significantly, so that search for new solutions includes a large random component to achieve suitably wide exploration. This search can be accompanied by a random generation and selection of new functional systems, similar to neural group selection in the theory of neural Darwinism [19]. Thus, the ordinary regime of operation can be considered as fine tuning of the animat control system, whereas the extraordinary regime is a coarse search for a suitable adaptive behavior in unexpected situations.

Our control system clearly assumes priority and supervision by high-level FS. For example, if a danger signal is received by FS1 (which controls security of the animat), while the animat is searching for food within the branch controlled by FS2 (corresponding to energy replenishment), the FS1 is able to interrupt the search for food and turn on a security behavior by activating an appropriate FS.

The memory of neural networks in the abovementioned FS may be prone to forgetting because of learning (reflecting the well known stability-plasticity dilemma). It is possible to include long-term memory of the acquired skills into our control system. If a certain type of behavior was well tested and became reliable, the corresponding FS could be copied into long-term memory, namely into the long-term FS, in which values ε and α_C , α_M are equal to zero. The both short-term FS (with normal values of ε and α_C , α_M or with large values of ε) and long-term FS could perform the same operations in the same design structure. For reliable skills, the long-term FS have priority over the short-term FS. However, if prognoses of states $\mathbf{S}^{\mathbf{pr}}$ made by the long-term FS differ from actual states $\mathbf{S}$, the control returns to the short-term FS.

V. Discussion and Conclusion

The proposed ACD based scheme of the animat control system provides a general approach to modeling adaptive behavior of an animat with natural needs and corresponding motivations, goals and sub-goals. In particular, our proposal appears to be quite appropriate for developing Alife models of purposeful adaptive behavior of agents described in [20, 21] in which simple hierarchical structures of agent control systems are shown to have emerged through evolutionary self-organization of neural networks.

It should be noted that the system proposed here differs from the system described in our previous work [11], in which we assume that super-systems deliver more explicit tasks and goals to sub-systems. The current ACD based design seems to be more flexible than the design of [11]. However, the referenced architecture with explicit goals is still interesting and it can be used for further development of animat brain architectures.

The described version of the animat control system is based on a simplified version of ACD. Naturally, ACD architectures can include additional blocks (controller network, reference model, etc.) represented by more powerful recurrent neural networks [22], and provide opportunities to develop more intelligent animat brain designs. In particular, a reference model in ACD can operate analogously to the reference signal in Powers's perceptual control systems [23].

ACKNOWLEDGMENTS

The authors thank Prof. Konstantin V. Anokhin for discussing this manuscript, providing valuable comments and a description of early works by Petr K. Anokhin (Refs. [14-16]). This work is supported by the Russian Foundation for Basic Research, Grant No 04-01-00179 and the Program of the Presidium of the Russian Academy of Science "Mathematical Modeling and Intelligent Systems", Project No 16.2.45.

REFERENCES

[1] J.-A. Meyer, S.W. Wilson, (Eds.), *From Animals to Animats: Proceedings of The First International Conference on Simulation of Adaptive Behavior*. Cambridge, MA: The MIT Press/Bradford Books, 1991.

[2] S.W. Wilson, "The animat path to AI," In: [1], pp. 15-21.

[3] J.Y. Donnart and J.A. Meyer "Learning reactive and planning rules in a motivationally autonomous animat," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 26, no.3, pp. 381-395, 1996.

[4] P.K. Anokhin, *Biology and Neurophysiology of the Conditioned Reflex and Its Role in Adaptive Behavior*, Pergamon, Oxford, 1974.

[5] B. Widrow, N. Gupta, and S. Maitra, "Punish/reward: learning with a critic in adaptive threshold systems," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no.5, pp. 455-465, 1973.

[6] P.J. Werbos, "Advanced forecasting methods for global crisis warning and models of intelligence," *General Systems Yearbook*, vol. 22, pp.25-38, 1977.

[7] A.G. Barto, R.S. Sutton, and C.W. Anderson, "Neuronlike elements that can solve difficult learning control problems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 13, pp. 835-846, 1983.

[8] P.J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," In: *Handbook of Intelligent Control,* White and Sofge, Eds., Van Nostrand Reinhold, pp. 493 – 525, 1992.

[9] D.V. Prokhorov and D.C. Wunsch, "Adaptive critic designs," *IEEE Trans. Neural Networks*, vol. 8, no.5, pp. 997-1007, 1997.

[10] R.S. Sutton, "Reinforcement learning architectures for animats," In: [1], pp. 288-296.

[11] K.V. Anokhin, M.S.Burtsev, I.Yu. Zarayskaya, A.O. Lukashev, V.G. Red'ko, "Project "Animat Brain": designing model of adaptive behavior on the base of functional system theory," *Eighth national conference on artificial intelligence. Proceedings*, Moscow: Fizmatlit, vol. 2, pp. 781-789, 2002. (In Russian)

[12] M.M. Bongard, I. S. Losev, M.S. Smirnov, "Project of a model of organization of behavior – "Animal"," In: *Modeling of Learning and Behavior*, Moscow: Nauka, 1975. pp. 152-171. (In Russian).

[13] P.J. Werbos, "A brain-like design to learn optimal decision strategies in complex environments," In: *Brain-like Computing and Intelligent Information Systems*, Ed. Shun-ichi Amari, Nikola Kasabov, Springer, pp. 264-303, 1998.

[14] P.K. Anokhin, "New data on the problem of interactions between center and periphery in the neural activity," *Archive of Biol. Sci.*, vol.12, pp. 290-338, 1937 (in Russian).

[15] P.K. Anokhin, *Problem of Center and Periphery in the Physiology of Neural Functions*, Gorky: Medical Publishing House, 1935 (in Russian).

[16] P.K. Anokhin, "Key problems in the studies of the higher nervous functions," In: *Problems of the Higher Nervous Functions*. Ed. P.K.Anokhin. Moscow. Academy of Medical Sci. Press. pp. 9-128, 1949 (In Russian).

[17] E.A. Umryukhin, *Brain Mechanisms: Informational Model and Optimization of Learning*. Moscow, 1999 (In Russian).

[18] R.S. Sutton, A.G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA: A Bradford Book, 1998.

[19] G. M. Edelman, *Neural Darwinism: The Theory of Neuronal Group Selection*, Oxford: Oxford University Press, 1989.

[20] M.S. Burtsev, "The model of emergence of goal-directed adaptive behavior. II. Evolution of hierarchy of goals," Institute of Applied Mathematics of RAS, Preprint no. 69, 2002 (In Russian).

[21] M.S. Burtsev, "Measuring the dynamics of artificial evolution," In: W. Banzhaf et al. (Eds.), *European Conference on Artificial Life*, LNAI 2801, pp. 580-587, 2003.

[22] D.V. Prokhorov, Backpropagation through time and derivative adaptive critics: a common framework for comparison. In Si et al. (Eds.), *Learning and Approximate Dynamic Programming*, IEEE Press, 2004 (in press).

[23] W.T. Powers, *Behavior: the Control of Perception*, Aldine, Chicago, 1973.