Project "Animat Brain": Designing the Animat Control System on the Basis of the Functional Systems Theory

Vladimir G. Red'ko[1], Konstantin V. Anokhin[2], Mikhail S. Burtsev[3], Alexander I. Manolov[4], Oleg P. Mosalov[4], Valentin A. Nepomnyashchikh[5], Danil V. Prokhorov[6]

[1] Center of Optical Neural Technologies, Scientific-Research Institute for System Studies,
Russian Academy of Sciences,
Vavilova Str., 44/2, Moscow, 119333, Russia
vgredko@gmail.com

[2] P.K. Anokhin Research and Development Institute of Normal Physiology, Russian
Academy of Medical Sciences, Mokhovaya Str., 11/4, Moscow, 103009, Russia
k_anokhin@yahoo.com

[3] M.V. Keldysh Institute for Applied Mathematics, Russian Academy of Sciences,
Miusskaya Sq., 4, Moscow, 125047, Russia
mbur@narod.ru

[4] Moscow Institute of Physics and Technologies,
Institutsky per., 9, Dolgoprudny, Moscow region, 141700, Russia
olegmos_@mail.ru, paraslonic@yandex.ru

[5] I.D. Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences,
Borok, Yaroslavl region, 152742, Russia
nepom@ibiw.yaroslavl.ru

[6] Toyota Technical Center in Ann Arbor, MI, USA
dvprokhorov@gmail.com

Abstract. The paper proposes the framework for an animat control system (the Animat Brain) that is based on the Petr K. Anokhin's theory of functional systems. We propose the animat control system that consists of a set of functional systems (FSs) and enables predictive and purposeful behavior. Each FS consists of two neural networks: the actor and the predictor. The actors are intended to form chains of actions and the predictors are intended to make prognoses of future events. There are primary and secondary repertoires of behaviors: the primary repertoire is formed by evolution; the secondary repertoire is formed by means of learning. The paper describes both principles of the Animat Brain operation and the particular model of predictive behavior in cellular landmark environment.

Keywords: Animat control system, predictive behavior, learning, evolution.

1 Introduction

This paper proposes the framework for an animat control system (the Animat Brain) that is based on the biological theory of functional systems. This theory was proposed and developed in the period 1930-1970s by Russian neurophysiologist Petr K.

Anokhin [1] and pays special attention to prediction and anticipation of a final needful result of a goal-directed action.

There are a number of researches that analyze prediction and anticipation in animat control systems [2,3]. Tani investigated recurrent neural network (RNN) approach implementing predictive models for mobile robots [4,5]. Witkowski proposed the expectancy model that is based on a set of heuristic rules [6]. Butz et al [7] developed anticipatory learning classifier systems (ALCSs) that incorporate methods of reinforcement learning, genetic algorithm and earlier versions of classifier systems [8,9].

The main goal of our work is to propose the neural network (NN) animat control system that enables explicit models of predicted states. The architecture of the NN control system is formed by biologically plausible self-organizing processes. We also propose simple cellular environments that can be used in both biological and computer simulation experiments.

The ideas of our work are similar to that developed in Tani's research [4,5], however we propose more distributed NN architecture as compared with RNN. The explicit NN models of predicted states in our approach are similar to SAS (Sign-Action-Sign) relations in Witkowski's dynamic expectancy model [6]. A more detailed comparison of our approach with other works will be given at the end of the paper.

The paper is organized as follows. Section 2 outlines Anokhin's theory of functional systems. Section 3 describes principles of animat control system operation. A particular example of the proposed model is described in Section 4. Section 5 contains discussion and conclusion.

2 Anokhin's Theory of Functional Systems

Functional systems were put forward by Petr K. Anokhin in the 1930s as an alternative to the predominant concept of reflexes [1]. Contrary to reflexes, the endpoints of functional systems are not actions themselves but adaptive results of these actions. According to the functional systems theory, initiation of each behavior is preceded by the stage of afferent synthesis (Fig.1). It involves integration of neural information from a) dominant motivation (e.g., hunger), b) environment (including contextual and conditioned stimuli), and c) memory (including innate knowledge and individual experience). The afferent synthesis ends with decision making, which results in selection of a particular program of an action.

A specific neural module, acceptor of the action result, is being formed before the action itself. The acceptor stores an anticipatory model of the needful result of a goal-directed action. Such a model is based on a distributed neural assembly that includes various parameters (i.e., proprioreceptive, visual, auditory, olfactory) of the expected result. Execution of every action is accompanied by a 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, a new functional system is formed and all operations of the functional system are repeated. Such processes take place until the final needful result is achieved.

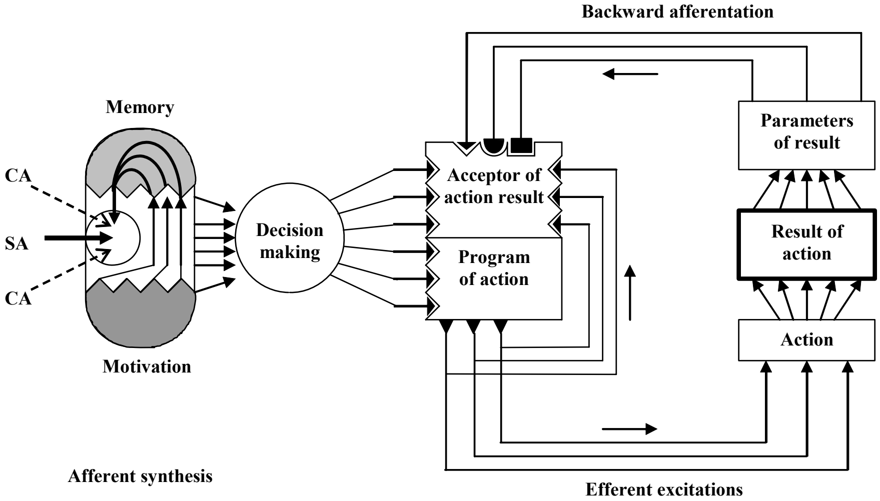

Fig. 1. General architecture of a functional system. SA is starting afferentation, CA is contextual afferentation. Operation of the functional system includes: 1) preparation for decision making (afferent synthesis), 2) decision making (formation of a program of an action), 3) prognosis of the action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

A separate 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 two of these mechanisms: evolution and learning.

3 Architecture and Principles of Operation of the Animat Brain

It is supposed that the animat control system consists of neural network (NN) blocks and is analogous to an animal control system. Each block is a formal functional system (FS). At any moment in time (t = 1,2,…), only one FS is active, in which the current action is formed. There are connections between FSs; the active FS can transmit activation to every FS through these connections.

Each FS consists of two NNs: the actor and the predictor. Operation of the active FS can be described as follows. The state vector $\mathbf{S}(t)$ characterizing the current external and internal environment is fed to the FS input. The actor forms the action $\mathbf{A}(t)$ in accordance with given state $\mathbf{S}(t)$, i.e. the actor forms the mapping $\mathbf{S}(t)$ -> $\mathbf{A}(t)$. The predictor makes prognosis of the next state for given vectors $\mathbf{S}(t)$ and $\mathbf{A}(t)$, i.e. the predictor forms the mapping {$\mathbf{S}(t)$, $\mathbf{A}(t)$} -> $\mathbf{S}^{\mathbf{pr}}(t+1)$. So, the predictor stores a model of casual relation between the current state $\mathbf{S}(t)$, action $\mathbf{A}(t)$ and the next state $\mathbf{S}(t+1)$. The prediction $\mathbf{S}^{\mathbf{pr}}(t+1)$ of the next state corresponds to the acceptor of action result in the functional system theory. The mappings $\mathbf{S}(t)$ -> $\mathbf{A}(t)$ and {$\mathbf{S}(t)$, $\mathbf{A}(t)$} -> $\mathbf{S}^{\mathbf{pr}}(t+1)$ are stored in NN synaptic weights.

Activation is transmitted from one FS to others in accordance with connectivity matrix C_{ij} , the value C_{ij} characterizes the probability that the *j*-th FS is activated by the *i*-th FS.

The animat receives reinforcements (rewards and punishments) which are related to animat needs.

It is supposed that there are primary and secondary repertoires of behaviors. The primary repertoire is formed by evolution: there is a population of animats and a set of FSs, synaptic weights of NNs and connectivity matrix C_{ij} are adjusted during evolutionary processes.

The secondary repertoire of behavior is formed by learning. There are two regimes of learning: 1) the extraordinary mode and 2) the fine tuning mode.

The extraordinary mode is a rough search of behavior that is adequate to the current situation. This mode comes, if the predicted state $\mathbf{S}^{\mathbf{pr}}(t+1)$ in the active FS strongly differs from the real state $\mathbf{S}(t+1)$. In terms of the functional system theory, large difference between $\mathbf{S}^{\mathbf{pr}}(t+1)$ and $\mathbf{S}(t+1)$ means that parameters of the result differ essentially from parameters stored in the acceptor of action result.

In the extraordinary mode, a random search for new behaviors takes place; namely, the connectivity matrix C_{ij} is substantially changed, new FSs can be randomly generated and selected. This mode is similar to neural group selection in Edelman's theory of Neural Darwinism [10].

In the fine tuning mode, learning is adjustment of NN weights in the FS that is active at the current moment of time and in the FSs that were active in several previous steps of time. As synaptic weights are updated in those NNs, which were active in previous time steps, this learning mode allows forming chains of consecutive actions. Synaptic weights in predictors are modified to minimize prediction errors (e.g. by means of error back-propagation [11]). Synaptic weights in actors are adjusted by a Hebbian-like rule: the synaptic weights in actors are modified to make the mappings $\mathbf{S}(t)$ -> $\mathbf{A}(t)$ stronger/weaker for positive/negative reinforcements.

We introduce two modes of learning (the extraordinary and the fine tuning mode) for the following reasons:

1) We believe that learning by means of these two modes (rough search in a quite new situation and fine tuning in a partially known situation) is more effective as compared with one mode.

2) Analogies to the fine tuning and extraordinary modes can be seen in animal behavior. For example, bees, butterflies and other insects are able to crop pollen or nectar from various flower species, but individual insects tend to choose flowers of a particular species, while ignoring others. The reason for this so called "flower constancy" is that different flower species differ by their structure, so insects should learn a structure of particular flowers to extract food efficiently. The learning takes a number of visits to the same flower species, resulting in a gradual decline in handling time on successive visits [12,13]. This learning is analogous to our fine tuning mode. If a production of food by preferred flower species falls low, then an insect starts to sample various other species [12] (an extraordinary mode).

An extraordinary mode also can be seen under artificial conditions unfamiliar to an animal. For example, certain species of jumping spiders live on trees and have no experience with water spaces. When faced a water-filled tray in laboratory for the first time, they may choose one of only two solutions available for them: jump over the tray or “swim” (in fact, walk) across it. Individuals, which attempt to swim first and fail to reach the opposite side of the tray, switch to jumping if allowed a next trial. Those animals, which jump first and fail, switch to swimming (the extraordinary mode). If a spider reaches the opposite side, it repeats a successful behavior (swimming or jumping) in the course of next trials. Once an appropriate behavior is chosen, only minute quantitative details of this behavior are varied [14]. One may speculate that these minute variations help to improve spider’s performance (the fine tuning mode of learning).

Existence of extraordinary and fine tuning modes in various animal species suggests that a learning based on two very different modes could be of adaptive value.

A particular version of the Animat Brain model is described in the next section. It should be underlined that we propose only possible version of the model. In order to ensure that all components of the model are consistent with each others, we describe concrete possible mechanisms of NNs operation, leaning, and evolution. In the current work we consider simplest variants of these mechanisms; in further research these mechanisms could be replaced by similar ones. We propose also certain landmark environment that can be used to compare behavior of simulated and real animals in the same model “world”.

4 Particular Model of Animat Brain Operation

4.1 Animat Environment and Features

Environment. We assume simple 2D cellular landmark environment (Figs. 2,3). Any marked cell A, B, C, D, G has its own landmark. The modeled “world” is restricted by impenetrable barriers. The animat sensory system is able to perceive the state of a marked cell (5 different signals), an unmarked cell and a cell of the barrier. So, there are 7 different possible signals from cells. The goal cell is G.

Animat Features. An animat senses its local environment and executes some actions. Actions are executed in accordance with the commands of the active FS of the animat control system. At any moment in time, the animat executes one of the following five actions:
1- 4) to move on one cell up/down/right/left,
5) to wait.

The animat has internal energy resource R. Performing actions, an animat spends its resource. We suppose that at every movement (actions 1-4), the animat resource is decreased by r_1, and when waiting (action 5), the decrease of the resource is negligible. Reaching the goal cell G, the animat increases its resource by r_2.

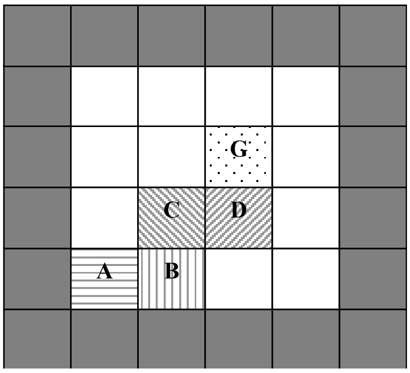

Fig. 2. Simple cellular environment. The landmarks A, B, C, D, G are in adjacent cells. The goal cell is G. The "world" consists of 4x4 cells; it is surrounded by impenetrable barriers (grey cells)

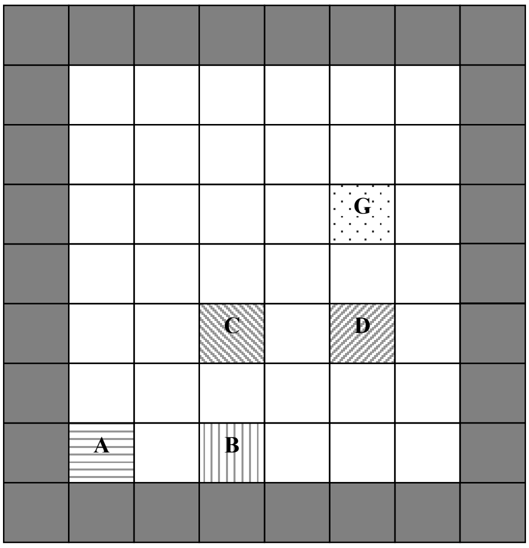

Fig. 3. The cellular environment that is similar to the "world" in Fig. 2, but the landmarks A, B, C, D, G are separated by one cell distance

Animat Sensory System. The animat perceives states of five cells: its own cell and four cells around (up/down/right/left). In each of four surrounding cells, the animat estimates one of 7 signals (5 kinds of landmarks, unmarked cell or barrier cell); in its own cell, the animat estimates one of 6 signals (5 kinds of landmarks or unmarked cell). For definiteness we suppose that every such signal is a binary component (1 or 0) of the state vector **S**(t). Also the animat perceives its resource $R(t)$ and the resource change for last time step $R(t)$ - $R(t\text{-}1)$. As other signals take values 0 or 1, it is convenient to characterize resource and resource change by binary values too. So, we assume that the animat estimates the binary values S_R and S_{DR} , where S_R = 1 if $R(t) > R_T$ and $S_R = 0$ if $R(t) < R_T$ (R_T is a predetermined threshold resource value), and

S_{DR} =1 if $R(t) > R(t\text{-}1)$ and S_{DR} = 0 if $R(t) < R(t\text{-}1)$. Thus, the animat perceives 36 binary parameters, characterizing its external and internal environment. These 36 values form the current state vector $\mathbf{S}(t)$.

4.2 Animat Control System

The animat control system is a set of FSs; each FS consists of two neural networks: the actor and the predictor. At any moment in time, only one FS is active, in which the current action is formed; after performing its operation, this FS transmits activation to other FS. The new active FS is chosen probabilistically. The probability that the j-th FS is activated by the i-th FS is equal to $C_{ij} /(\sum_k C_{ik})$, with C_{ij} as the element of the connectivity matrix ($C_{ij} > 0$).

Neural Network of the Actor. The actor is a two layer NN. The operation of the actor is described by the following equations:

$$\mathbf{x}^\mathbf{A} = \mathbf{S}(t) , \quad y^A_j = \mathrm{th}\,(\sum_i w^A_{ij} x^A_i) , \quad z^A_k(t) = \mathrm{F}(\sum_j v^A_{jk} y^A_j) , \qquad \mathbf{(1)}$$

$$\mathrm{F}(a) = 1/[1+\exp(-a/b)] , \qquad \mathbf{(1a)}$$

where $\mathbf{x}^\mathbf{A}$ is the NN input vector, it is equal to the current state vector $\mathbf{S}(t)$, $\mathbf{y}^\mathbf{A}$ is the vector of hidden layer outputs, $z^A_k(t)$ are signals of output layer neurons, w^A_{ij} and v^A_{jk} are NN synaptic weights, F(.) is the sigmoid activation function, parameter b regulates the slope of this function. The probability that m-th action is selected is equal to $z^A_m(t) \,/\, \sum_k z^A_k(t)$. The action vector $\mathbf{A}(t)$ is determined as follows: $A_m(t)$ = 1, if m-th action is selected, all other components of $\mathbf{A}(t)$ are set to be equal 0.

Neural Network of the Predictor. The predictor is also a two layer NN. The operation of the predictor is described by the following equations:

$$\mathbf{x}^\mathbf{P} = \{\mathbf{S}(t), \mathbf{A}(t)\}, \quad y^P_j = \mathrm{th}(\sum_i w^P_{ij} x^P_i) , \quad z^P_k(t\text{+}1) = \mathrm{F}_1(\sum_j v^P_{jk} y^P_j) , \qquad \mathbf{(2)}$$

$$S^{pr}_k(t\text{+}1) = 1 \ \text{ if } \ z^P_k(t\text{+}1) > 0.5 , \quad S^{pr}_k(t\text{+}1) = 0 \ \text{ if } \ z^P_k(t\text{+}1) < 0.5 , \qquad \mathbf{(2a)}$$

$$\mathrm{F}_1(\mathrm{a}) = 1/[1+\exp(\text{-a})] , \qquad \mathbf{(2b)}$$

where $\mathbf{x}^\mathbf{P}$ is the NN input vector, it is the compound vector $\mathbf{x}^\mathbf{P} = \{\mathbf{S}(t), \mathbf{A}(t)\}$, $\mathbf{y}^\mathbf{P}$ is the vector of hidden layer outputs, w^P_{ij} and v^P_{jk} are NN synaptic weights, $z^P_k(t\text{+}1)$ are signals of output layer neurons, $S^{pr}_k(t\text{+}1)$ are components of the predicted state vector $\mathbf{S}^\mathbf{pr}(t\text{+}1)$.

Eqs. (1,2) describe the simple version of NNs operation that ensures both natural schemes of learning (see below) and binary components of vectors $\mathbf{A}(t)$ and $\mathbf{S}^\mathbf{pr}(t\text{+}1)$.

4.3 Learning Mechanism

There are two regimes of learning: 1) the extraordinary mode and 2) the fine tuning mode.

The extraordinary mode occurs, if there is a strong mismatch between the expected and real results: the predicted state $\mathbf{S}^{\mathbf{pr}}(t+1)$ in the active FS essentially differs from the real state $\mathbf{S}(t+1)$. A strong mismatch means the difference in essential components of vectors $\mathbf{S}^{\mathbf{pr}}(t+1)$ and $\mathbf{S}(t+1)$: for example, the increase of the animat resource was expected, but really the resource was reduced.

In order to define essential components, we introduce a mask for every block. The mask is the vector $\mathbf{M}$ of dimension 36; this vector has components that are equal to 0 or 1. The unit components of the vector $\mathbf{M}$ define essential components of the state vector $\mathbf{S}(t+1)$. Namely, the component $S_k(t+1)$ is determined as essential, if $M_k =1$. If $M_k = 0$, the component $S_k(t+1)$ is considered as inessential. The essential components determine, which causal relation between the current state $\mathbf{S}(t)$, current action $\mathbf{A}(t)$ and next state $\mathbf{S}(t+1)$ is checked by the given predictor.

In the current version of our model it is supposed that the extraordinary mode of learning occurs as follows: a) activation is returned back to the i-th FS, that activated the current j-th FS, b) the element of the connectivity matrix C_{ij} corresponding to the link between these two FSs is changed.

The change of connection value C_{ij} occurs as follows. First, this value C_{ij} strongly decreases in the next time moment $t+1$, at which the i-th FS repeats activation of other FS. At this time moment, the temporary value of connection C_{ij}^{Temp} is used, and then there is a return to the usual connection value C_{ij} :

$$C_{ij}^{Temp}(t+1) \ = K_1 \ C_{ij} \ (t) \ . \qquad \textbf{(3a)}$$

Secondly, the connection value C_{ij} slightly decreases in long-term manner:

$$C_{ij} \ (t+2) = K \ C_{ij} \ (t) \ , \qquad \textbf{(3b)}$$

where $0 < K_1 < K < 1$. For example, we can set $K_1 = 0.1$, $K = 0.9$.

The described scheme of adjusting the connection value C_{ij} suggests that activation is transferred with high probability from the j-th FS that performed “unsatisfactory” action at the moment t to some other FS and then the probability to activate the j-th FS in future is slightly reduced.

Learning in extraordinary mode means that there is certain reorganization of animat control system operation. It is also possible to implement random generation and selection of new FSs in the extraordinary mode of learning; we intend to consider this option in further versions of the Animat Brain.

During *the fine tuning mode*, learning occurs by adjusting NN synaptic weights. This learning takes place when there is no strong mismatch between the expected and obtained result. Learning in actors and predictors occurs in different ways.

Learning in actors occurs according to reinforcements. Synaptic weights are adjusted in the FS that is active at the current moment of time *t*, and in FSs, that were active several previous steps of time. These synaptic weights are modified as follows:

$$\Box W_{ij} = \Box_A \Box^k X_i(t\text{-}k)\ Y_j(t\text{-}k)\ [R(t) - R(t\text{-}1)]\ , \qquad \textbf{(4)}$$

where W_{ij} is the weight of the considered synapse, $X_i(t\text{-}k)$ is the signal on the synapse input, $Y_j(t\text{-}k)$ is the output of the neuron corresponding to the given synapse, $\Box_A$ is learning rate of actors, $\Box$ is the discount factor ($0 < \Box < 1$), *k* is the difference between the current moment of time and the time of operation of the considered FS, $[R(t) - R(t\text{-}1)]$ is the value of the current reinforcement.

As learning occurs in those actors, which were active in several previous steps of time, this type of training allows forming chains of actions.

Learning in the predictor occurs, if there is a mismatch between the prediction $\mathbf{S}^{\mathbf{pr}}(t+1)$ and the result $\mathbf{S}(t+1)$ in any components of these vectors.

Learning in the predictor is carried out by the usual method of error back-propagation [11]. At this learning the target vector is $\mathbf{S}(t+1)$, and the NN output vector (that is compared with the target vector) is the vector $\mathbf{z}^{\mathbf{P}}(t+1)$, that is formed at the output layer of the predictor NN, see formulas (2).

In addition to fine tuning mode, we consider *learning upon achievement of the final needful result* (see description of the functional system theory in Section 2). We suppose that, upon achievement of the final needful result, there is strengthening connections between several FSs, which were active immediately before achievement of this result. In the current model the final needful result corresponds to reaching the goal cell G. For this type of learning connections between FSs are modified as follows:

$$\Box\Box_{ij} = \Box_L\ (\Box_L)^k\ r_2\ \ , \qquad \textbf{(5)}$$

where $\Box_{ij}$ is the connection between considered FSs, $\Box_L$ and $\Box_L$ are learning rate and the discount factor for this type of learning, *k* is the difference between the reward time and the time of considered activation transfer, r_2 is the value of the reinforcement in the cell G.

4.4 Evolution Mechanism

We consider a simple genetic algorithm (GA) [15,16] that can be described as follows. An evolving population consists of *n* animats. Evolution passes through a number of generations, $n_g = 1,2,\ldots$ At any generation, each animat is tested during *T* time steps independently of other animats of the population. At the beginning of the test, the animat resource $R(t)$ is set to certain predetermined value R_0 and the animat itself is set into the cell A. Then the animat acts in accordance with its control system and its resource is changed according to reinforcements. When the animat reaches the goal cell G and receive the reward r_2, it is returned to the start cell A. Such process is repeated, until the time T is over. After testing all n animats, the transition to the new generation occurs. At this moment, the animat having the maximum resource $R_{max}(n_g)$ is determined. This best animat gives birth to n children that constitute a new (n_g+1)-th generation.

The initial architecture of the animat control system (the set of FSs and the connectivity matrix C_{ij}) as well as initial synaptic weights of NNs form the animat genome **G**. The genome **G** is received at animat birth and is not changed during animat life. It is transferred (with small mutations) from the parent (the best animat of n_g-th generation) to descendants (all animats of (n_g+1)-th generation). Temporary architecture and synaptic weights of the NNs are changed during animat life via learning described in section 4.3.

At the beginning of (n_g+1)-th generation, the genome **G** of each newborn animat is determined: the offspring genomes are obtained from the genome of the parent through mutations that include:

1) duplication (with certain probability P_D) of every existing FS;
2) forming of elements of the connectivity matrix C_{ij} , corresponding to new FSs;
3) removing (with certain probability P_R) of every existing FS;
4) small random variations of elements of the connectivity matrix C_{ij} and synaptic weights of all NNs;
5) small random variations of the mask vector **M** for every predictor.

The described evolution mechanism is the simple version of the GA that takes into account all compounds of the current Animat Brain model. Similar and more sophisticated versions of the GA [15,16] could be used in future research.

4.5 Interaction between Selection of Actions and Predictions

In the current model, we pay special attention to predictions of future states. We suppose that essential learning takes place in the extraordinary mode, when there is large difference between predictions and results of action. This implies that chains of actions (formed by actors) should correspond to predictions (formed by predictors).

For example, consider the “world” shown in Fig. 3. When the animat placed in the cell A moves two times right, it should be able to predict the movement into an unmarked cell after the first step and into the landmark cell B after the second step. Moving further two times upwards, it should predict the displacement into an unmarked cell and into the landmark cell C after the first and second steps, respectively. Then it should be able to predict movements to the landmark cells D and G. In principle, the animat can find an alternative path to the goal cell G, however, using landmarks, it is able to find the reliable path. Chains of actions and predictions should be in agreement with each other for reliable behavior.

Thus, we plan to analyze, how the agreement between chains of actions and predictions can be formed through learning and evolution in the current model.

5 Discussion and Conclusion

Comparison with Other Approaches. As was stated in the Introduction, our approach is similar to models by Tani [4,5], who models predictive behavior of mobile robots using RNNs. As compared with Tani's works, our model provides more explicit representation of states $\mathbf{S}(t)$, actions $\mathbf{A}(t)$ and predictions $\mathbf{S}^{\mathbf{pr}}(t+1)$.

Referring to Witkowski work [6] and research by Butz et al [7], we can note that our NN approach is based on the biological theory of functional systems [1] and we believe that it will be more flexible as compared with rule-based methods used in [6,7].

We can also compare our approach with works by Edelman et al, who investigate adaptive behavior that is controlled by huge NN control systems [17,18]. Our approach is at intermediate positions between small NN control system investigated in [4,5] and very large NN "brains" simulated in [17,18].

Our model includes two types of learning: 1) the extraordinary mode and 2) the fine tuning mode, and this can provide additional advantages as compared with similar models [4-7,17,18].

In our previous work, we designed Animat Brain architecture that is based on the reinforcement learning (RL) and consists of a set of hierarchically linked FSs [19]. Every FS is a simple adaptive critic design (ACD) that consists of two NNs: the model (predictor) and the critic; the model is intended to predict the next state $\mathbf{S}(t+1)$ for given current state $\mathbf{S}(t)$ and all possible actions a_i (the number of actions a_i is supposed to be small); the critic is intended to estimate state value function $V(\mathbf{S}(t))$. Actions are chosen in accordance with ε-greedy rule [8] ensuring selection of those actions that maximize state values V. However, analyzing evolution and learning in populations of such adaptive critics [20], we observed that ACD operation can be evolutionary unstable. This is due to necessity to estimate state value function $V(\mathbf{S})$; these estimations impose too strong a restriction on adaptive agent functioning. In the current work we introduce Hebbian-like learning modulated by rewards and punishments instead of the usual RL scheme. Similar viewpoint on RL and evolution was expressed by Stanley, Bryant and Miikkulainen [21], who emphasized that discovering complex NN control systems of adaptive agents by means of evolution is more effective than RL. In contrast to neuroevolution method [21], our schemes of search for adaptive behavior by both evolution and learning correspond to the biologically inspired concept of primary (formed by evolution) and secondary (formed by learning) repertoire of behaviors.

Our approach is similar to works by Wolpert, Kawato et al [22,23] on multi-modular NN systems for motor control. The architectures investigated in [22,23] include multiple pairs of inverse (controller) and forward (predictor) models. The inverse model is similar to the actor in our architecture; the forward model plays the same role as the predictor in our schemes. It should be noted that we consider the control system of an autonomous animat, whereas Wolpert, Kawato et al analyze learning at human motor control that corresponds to psychological experiments on movement of different objects at different conditions by an arm.

It should be underlined than simulation of adaptive behavior in landmark "worlds" proposed in this work (Figs. 2,3) can be used for comparison of different approaches, such as RNN [4,5], adaptive critic designs [19], brain-inspired NN control system [17,18], ALCS [7], and distributed NN-based FSs.

Possible Variations on the Proposed Model. One of the difficulties of the current model is too large dimension of state vectors $\mathbf{S}(t)$ that include 36 components. To overcome this difficulty we can consider more specialized FSs. A particular FS can perceive only a small subset of parameters from the local environment. For example, the FS that is responsible for movement from the cell A to the cell B (Fig. 2) can perceive only landmarks A and B and only in left and right cells. Such specialization can be implemented by means of mask vectors $\mathbf{M}^*$ that have components 0 or 1. Parameters corresponding to zeros (M^*_k =0) are not included into state vectors for the considered FS. This option can provide a distributed animat control system, in which many small specialized FSs constitute the whole Animat Brain. The specialized FSs can be formed through evolution and extraordinary mode of learning. It should be noted that this scheme of small specialized FSs is similar to the multi-modular architecture that was proposed and investigated in [22,23]. We can also consider the concept of module responsibility from [22,23] in order to organize a flow of FS activity throughout the Animat Brain architecture.

Figs. 2,3 show simple landmark “worlds”. Obvious generalizations and variations are possible: several different goals can be introduced; the landmark distribution can be unstable, noisy, etc.

Biological Aspects. We propose to investigate animat behavior in landmark environments (simple examples of which are shown in Figs. 2,3). This is interesting from a biological viewpoint for the following reasons:

- It is possible to design the cellular “world” with exactly the same structure for real biological experiments. Namely, we can construct the 2D array of cells with nontransparent walls between cells, color floor in certain cells by different landmarks and make a door between every neighboring pair of cells. Any door is automatically closing but it can be opened by an investigated animal.
- Landmarks are really used by animals in adaptive behavior. For example, honey bees use landmarks for efficient goal navigation [24].
- In some biological experiments, such as investigations of rat orientation in a Morris water maze [25], animals seem to be able to select and use landmarks to find a goal.

So, we can state that it is possible to compare the goal-directed behavior of simulated and real animals in proposed landmark environments.

Conclusion. We proposed the biologically inspired Animat Brain architecture that consists of a set of functional systems (FSs). Every FS includes two NNs: the actor and the predictor and provides action selection and predictions of action results. In the case of unexpected events, considerable learning takes place and animat behavior is reorganized. We intend to study conditions for which predictions of future events (formed by predictors) and generations of action chains (formed by actors) are consistent with each other. We also propose to investigate the predictive animat behavior in landmark environments that ensure comparison of behavior of simulated and real animals in the same model "world".

Acknowledgments. This work is supported in part by the Russian Foundation for Basic Research, Grants No 05-07-90049 and No 04-01-00510, the Program No 14 of the Presidium of the Russian Academy of Science on "Intelligent Technologies and Mathematical Modeling", and "Russian Science Support Foundation". The authors thank Martin Butz, Stewart Wilson and anonymous reviewers for valuable comments on the earlier versions of this paper.

References

1. Anokhin, P.K.: Biology and Neurophysiology of the Conditioned Reflex and Its Role in Adaptive Behavior. Pergamon, Oxford (1974)
2. Butz, M. V., Sigaud, O., Gérard, P. (eds.): Anticipatory Behavior in Adaptive Learning Systems. Springer-Verlag, Berlin (2003)
3. Butz, M.V., Sigaud, O., Gérard, P.: Internal Models and Anticipations in Adaptive Learning Systems. In [2], 86-109
4. Tani, J.: Model-Based Learning for Mobile Robot Navigation from the Dynamical Systems Perspective. IEEE Trans. on Systems, Man, and Cybernetics. Part B: Cybernetics. 26 (1996) 421-436
5. Paine, R W., Tani, J.: How Hierarchical Control Self-organizes in Artificial Adaptive Systems. Adaptive Behavior.13 (2005) 211-225
6. Witkowski, M.: Towards a Four Factor Theory of Anticipatory Learning. In [2], 66-85
7. Butz, M.V., Goldberg, D. E.: Generalized State Values in an Anticipatory Learning Classifier System. In [2], 282-302
8. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA: A Bradford Book (1998)
9. Holland, J.H., Holyoak, K.J., Nisbett, R.E., Thagard, P.: Induction: Processes of Inference, Learning, and Discovery. Cambridge, MA: MIT Press (1986)
10. Edelman, G.M.: Neural Darwinism: The Theory of Neuronal Group Selection. Oxford University Press, Oxford (1989)
11. Rumelhart, D.E., Hinton, G.E., Williams, R.G.: Learning Representation by Back-Propagating Error. Nature. 323 (1986) 533-536
12. Chittka, L., Thomson, J.D., Waser, N.M.: Flower Constancy, Insect Psychology, and Plant Evolution. Naturwissenschaften. 86 (1999) 361–377
13. Goulson, D., Stout, J.C., Hawson S.A.: Can Flower Constancy in Nectaring Butterflies Be Explained by Darwin's Interference Hypothesis? Oecologia. 112 (1997) 225–231
14. Jackson, R. R., Carter, C. M., Tarsitano, M. S.: Trial-and-error Solving of Confinement Problem by Araneophagic Jumping Spiders, Portia Fimbriata. Behaviour. 138 (2001) 1215-1234

15. Holland, J. H.: Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor, MI. (1975). 2nd edn., MIT Press, Boston, MA (1992)
16. Goldberg, D. E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley (1989)
17. Krichmar, J.L., Edelman, G.M.: Machine Psychology: Autonomous Behavior, Perceptual Categorization and Conditioning in a Brain-Based Device. Cerebral Cortex. 12 (2002) 818-830
18. Krichmar, J.L., Seth, A.K., Nitz, D.A., Fleischer, J.G., Edelman, G.M.: Spatial Navigation and Causal Analysis in a Brain-Based Device Modeling Cortical-Hippocampal Interactions. Neuroinformatics. 3 (2005) 197-221
19. Red'ko, V.G., Prokhorov, D.V., Burtsev, M.S.: Theory of Functional Systems, Adaptive Critics and Neural Networks. In Proc. International Joint Conference on Neural Networks (IJCNN 2004), Budapest (2004) 1787-1792
20. Red’ko, V.G., Mosalov, O.P., Prokhorov, D.V.: A Model of Evolution and Learning. Neural Networks. 18 (2005) 738-745
21. Stanley, K.O., Bryant, B.D., Miikkulainen, R.: Evolving Neural Network Agents in the NERO Video Game. In Proceedings of the IEEE 2005 Symposium on Computational Intelligence and Games (CIG’05), Essex University, Colchester, Essex, UK (2005) 182-189
22. Wolpert, D.M., Kawato, M.: Multiple Paired Forward and Inverse Models for Motor Control. Neural Networks. 11 (1998) 1317-1329
23. Haruno, M., Wolpert, D.M., Kawato, M.: MOSAIC Model for Sensorimotor Learning and Control. Neural Computation. 13 (2001) 2201-2220
24. Fry, S.N., Wehner, R.: Look and Turn: Landmark-Based Goal Navigation in Honey Bees. The Journal of Experimental Biology. 208 (2005) 3945-3955
25. Morris, R.G.M., Garrud, P, Rawlins, J.N.P., O’Keefe, J.: Place Navigation Impaired in Rats with Hippocampal Lesions. Nature. 297 (1982) 681-683