

УДК: 32.81

Поведение модельных организмов, обладающих естественными потребностями и мотивациями^a

©2012 Коваль А.Г.^{*1}, Редько В.Г.^{**2}

¹ *Национальный исследовательский ядерный университет «МИФИ», Москва, Россия*

² *Научно-исследовательский институт системных исследований, Российская академия наук, Москва, Россия*

Аннотация. Построена и исследована компьютерная модель автономных адаптивных агентов, обладающих естественными для живых организмов потребностями: питание, размножение, безопасность. В модели введены мотивации, соответствующие каждой из потребностей. При моделировании обнаружено, что агенты формируют цепочки действий. Продемонстрировано формирование циклов поведения, в которых последовательно удовлетворяются потребности питания, безопасности и размножения.

Ключевые слова: модельные организмы, потребности, мотивации, обучение с подкреплением, формирование цепочек действий.

1. ВВЕДЕНИЕ

Важными свойствами живых организмов являются их естественные потребности: потребность питания, потребность размножения, потребность безопасности. Хотя в последние годы активно ведутся исследования по изучению свойств автономных агентов (модельных организмов), см., например, обзоры [1,2], тем не менее, работы по моделированию процессов формирования адаптивного поведения агентов с естественными потребностями пока малочисленны.

В настоящей работе исследуется адаптивное поведение автономных агентов, обладающих потребностями, аналогичными основным потребностям биологических организмов. Также вводятся мотивации, каждая мотивация соответствует определенной потребности, и анализируется роль мотиваций в формировании поведения агентов.

В предыдущих работах было начато исследование автономных агентов, обладающих естественными потребностями. В [3] построена и исследована упрощенная компьютерная модель адаптивного поведения автономных агентов, имеющих указанные потребности. Каждый агент имел ресурс. В модели предполагалось, что мир, в котором находились агенты, состоял из двух клеток: одна клетка являлась опасной для агентов, вторая – безопасной, причем статус клеток периодически менялся: опасная клетка становилась безопасной, а безопасная, наоборот, становилась опасной. Агент, находящийся в опасной клетке, каждый такт времени терял большой ресурс. В мире была восполняемая пища агентов, агенты могли размножаться. При съедании пищи ресурс агента увеличивался, при выполнении действий – уменьшался. Система управления агента была основана на правилах вида «Если имеет место ситуация S , то необходимо выполнить действие A ». Правила модифицировались как путем обучения с подкреплением [4], так и в процессе

^aРабота выполнена при финансовой поддержке РФФИ, проект № 10-01-00129

*Anton.G.Koval@ya.ru

**vgredko@gmail.com

эволюционной оптимизации. Моделирование продемонстрировало формирование естественного поведения агентов. Также было показано, что при эволюционной оптимизации популяции агентов важную роль играет размножение.

В другой компьютерной модели [5,6] исследовалось поведение отдельного автономного агента, ведущего поиск пищи в двумерной клеточной среде. В части клеток были порции пищи. При съедании пищи ресурс агента пополнялся. Система управления агента также была основана на правилах вида $S \rightarrow A$, веса правил оптимизировались методом обучения с подкреплением. В результате обучения агент формировал обобщающие эвристики. Эти эвристики таковы: если имеется пища в той же клетке, в которой находится агент, то нужно съесть пищу; если в клетке агента нет пищи, но есть пища в клетке впереди или справа/слева от агента, то нужно выполнить действие «перемещение вперед» или «поворот направо/налево». Использование эвристик приводило к формированию последовательных цепочек действий, приводящих к пополнению ресурса агента. Дополнительно в компьютерную программу вводилась процедура усреднения: вычислялось среднее число применений данного действия для той или иной ситуации [6]. В результате усреднения агент формировал внутренние понятия «имеется пища в моей клетке», «имеется пища в клетке впереди меня», «имеется пища в клетке справа/слева от меня».

В модели [7,8] исследовалась динамика мотивации в случае, непосредственно связанном с биологическим экспериментом. Изучалась модель поискового поведения личинок ручейников, строящих чехол-домик из частиц разного размера и ведущих поиск скоплений подходящих частиц. Была введена мотивация к прикреплению частиц, которая регулировала инерционное переключение между поисковыми тактиками. Была показана адекватность модели биологическим экспериментальным данным.

Настоящая работа развивает изученные ранее модели [3,5–8]. Ниже излагается модель автономных адаптивных агентов, обладающих естественными потребностями (питание, размножение, безопасность) и соответствующими потребностям мотивациями.

2. МОДЕЛЬ АГЕНТОВ, ОБЛАДАЮЩИХ ЕСТЕСТВЕННЫМИ ПОТРЕБНОСТЯМИ И МОТИВАЦИЯМИ

Считаем, что агент обладает потребностями питания, безопасности и размножения, и соответствующими этим потребностям мотивациями (M_F , M_S , M_R). Дополнительно вводится мотивация к сохранению текущего состояния M_N (эта мотивация характеризуется ниже). Между потребностями вводится иерархия, при этом каждый такт времени одна из потребностей и соответствующая ей мотивация агента являются ведущими. Удовлетворение ведущей потребности является положительным подкреплением при обучении.

2.1. Модельный мир

Рассматривается популяция агентов. Каждый агент обладает внутренним ресурсом $R(t)$. Если ресурс агента уменьшается до нуля, то данный агент погибает. Время t дискретно. Каждый агент помещается в свою клетку, в этой же клетке имеется хищник, активность которого периодически меняется: через каждые T_p тактов времени активный хищник становится неактивным, неактивный – активным. У агента, находящегося в одной клетке с активным хищником, каждый такт времени ресурс $R(t)$ уменьшается на значительную величину r_{attack} . Так как разные агенты помещаются в разные клетки, то они ведут себя независимо друг от друга.

2.2. Система управления агента

Система управления агента основана на наборе правил вида: $S_k \rightarrow A_k$, где S_k – ситуация, A_k – действие, k – номер правила. Каждое правило имеет свой вес W_k , веса правил изначально случайны, а затем модифицируются методом обучения с подкреплением. Вектор ситуации S_k имеет три компоненты. Первая компонента определяет, активен ли хищник в клетке агента, эта компонента принимает два значения (0 либо 1). Вторая компонента равна номеру действия, выполненного агентом в предыдущий такт времени, эта компонента принимает шесть значений (от 1 до 6). Третья компонента соответствует ведущей мотивации и принимает четыре значения (по числу возможных мотиваций).

Агент может выполнять одно из следующих действий:

- 1) покой,
- 2) поиск пищи,
- 3) питание,
- 4) подготовка к размножению,
- 5) размножение,
- 6) оборона.

При выполнении агентом i -го действия ресурс агента уменьшается на величину r_i . При этом, если агент в текущий такт времени t выполнил действие «питание», а в предыдущий такт $t-1$ – действие «поиск пищи», то ресурс агента увеличивается на величину r_e . Если агент в текущий такт времени выполнил действие «размножение», а в предыдущий такт – действие «подготовка к размножению», и при этом ресурс агента $R(t)$ больше величины r_{minr} , то агент рождает потомка. Агент-родитель передает агенту-потомку значительную часть своего ресурса $R(t)$, равную $k_R R(t)$; $0 < k_R < 1$. При этом агент-потомок наследует веса правил W_k родителя. Если агент выполнил действие «оборона», то он в текущий и следующий такты времени защищён от активного хищника и потери ресурса r_{attack} . Когда ресурс агента становится меньше 0, этот агент погибает.

Всего возможно 48 ситуаций и 6 действий, которые формируют 288 правил. Каждый такт времени с вероятностью $1-\varepsilon$ выполняется то действие, для которого вес W_k соответствующего ему правила для текущей ситуации максимален, с вероятностью ε выполняется случайное действие.

2.3. Иерархия мотиваций

Мотивации основаны на удовлетворении потребности агента к поддержанию трёх факторов: насыщения (F_F), безопасности (F_S) и размножения (F_R). Считаем, что для каждого фактора имеется порог (T_F , T_S , T_R), при превышении фактором которого удовлетворяется соответствующая потребность.

Динамика факторов определяется следующим образом. Фактор насыщения пропорционален ресурсу агента:

$$F_F = k_C R(t) . \quad (1)$$

Фактор безопасности увеличивается при выполнении агентом действия «оборона» на ΔF_S и уменьшается на 1 при выполнении любого другого действия.

Фактор размножения увеличивается при успешном рождении потомка на ΔF_R , не меняется при выполнении действия «подготовка к размножению» и уменьшается на 1 при выполнении других действий.

Между мотивациями вводится следующая иерархия (перечисляем мотивации в порядке их приоритетности):

- 1) пищевая мотивация (наиболее приоритетна), M_F ,
- 2) мотивация безопасности, M_S ,

3) мотивация к размножению, M_R .

Для подтверждения именно такой иерархии был поставлен дополнительный компьютерный эксперимент: для каждой мотивации вводился коэффициент значимости (целое положительное число), и иерархия выстраивалась в соответствии с убыванием значимости. Проводилось моделирование согласно основной модели, только дополнительно при размножении значимости мотиваций передавались потомку со случайными изменениями. Для выявления эволюционно выгодной иерархии проводился расчет на большом промежутке времени, равном 500000 тактов. За это время сменилось примерно 360 поколений (это значит, что хотя бы один из агентов популяции после этого времени имел ~ 360 предшественников). После этих 500000 тактов в популяции было 98 агентов. Среди них у 91 агента наблюдалась описанная выше иерархия: (M_F, M_S, M_R) , а у 7 остальных агентов – немного модифицированная иерархия: (M_F, M_R, M_S) .

В основной модели считается, что мотивация присутствует, если соответствующий ей фактор меньше своего порога. Из присутствующих мотиваций ведущей является та, которая наиболее приоритетна в соответствии с иерархией. Если все факторы больше порога, то считается, что все потребности удовлетворены и ведущей становится дополнительная мотивация M_N , которую можно назвать мотивацией к сохранению состояния. Определение системой управления агента ведущей мотивации можно представить в виде иерархической схемы, представленной на рис. 1.

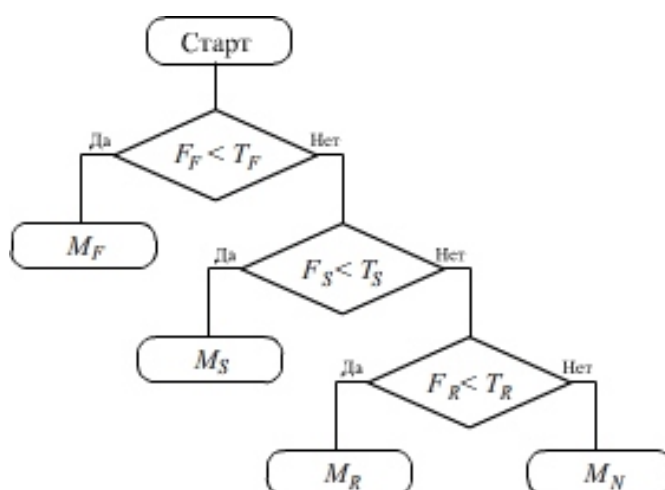


Рис. 1. Схема определения ведущей мотивации.

2.4. Схема обучения

В модели используется схема обучения с подкреплением [4]. Подкреплением является изменение фактора ведущей мотивации F_F , F_S или F_R :

$$\Delta W_{t-1} = \alpha (F_{L,t} - F_{L,t-1} + \gamma W_t - W_{t-1}), \quad (2)$$

где $F_{L,t}$, $F_{L,t-1}$ – факторы ведущей в такты t и $t-1$ мотивации, W_t и W_{t-1} – веса правил, примененных в такты t и $t-1$, α – параметр скорости обучения, γ – дисконтный фактор.

Изменение веса ΔW_{t-1} , примененного в такт времени $t-1$, происходит согласно (2) в момент времени t , а именно, тогда, когда уже становится известным фактор мотивации $F_{L,t}$ и вес выбираемого в этот момент правила W_t .

Если ведущей стала мотивация M_N , то агент выполняет действие «отдых», обучения при этом не происходит. Отметим, что при моделировании мотивация M_N становилась ведущей очень редко.

2.5. Результаты моделирования

При компьютерном моделировании параметры расчета были следующими: $\alpha = 0.1$, $\gamma = 0.8$, $\varepsilon = 0.05$, $k_C = 0.005$, $k_R = 0.25$, $T_F = 25$, $T_S = 25$, $T_R = 25$, $\Delta F_R = 50$, $\Delta F_S = 1$, $T_p = 10$, $r_e = 150$, $r_{minr} = 1000$, $r_{attack} = 100$, $r_1 = 1$, $r_2 = 5$, $r_3 = 5$, $r_4 = 10$, $r_5 = 10$, $r_6 = 10$.

Пороги и изменения факторов при выполнении действий выбирались так, чтобы поддержание пищевого фактора на уровне своего порога обеспечивало агенту достаточный запас ресурса для выживания после размножения (размножение – самое затратное по ресурсу действие, так как потомку передается 25% внутреннего ресурса родителя), достаточность определялась опытным путём; фактор размножения после действия «размножение» должен был намного превысить порог, так как в естественном поведении размножение происходит редко; поддержание фактора безопасности на уровне своего порога должно было обеспечивать достаточную частоту выполнения действия «оборона» в периодически возникающей опасной ситуации. Параметры, регулирующие изменения ресурса, выбирались так, чтобы поведение агента было достаточно естественным. Параметры обучения определялись опытным путём таким образом, чтобы обучение было эффективным.

В начале расчета (при $t = 0$) веса правил были случайны, за счет этого в некоторых компьютерных экспериментах популяция погибала, не успев обучиться. Такие эксперименты не брались в учёт.

При моделировании наблюдалось, что обучение приводило к формированию цепочек действий. Это регистрировалось следующим образом. Если записывать действия агентов как символы (6 символов для 6 различных действий), то на протяжении 2000 тактов времени получим строку из 2000 символов. Если символы распределены по строке случайно, то, в среднем, любая последовательность из 4 определённых символов повторяется $2000/(6^4) \approx 1.54$ раза. Если в реальной строке некоторая последовательность повторяется гораздо чаще, то можно считать, что это некий шаблон поведения. Для типичных расчетов был проанализирован ряд выполняемых действий агента с 3000 по 5000 такты времени, при этом были выделены 8 последовательностей из четырех действий, встречающихся чаще, чем 10 раз (до 30-кратного увеличения частоты встречаемости). Эти последовательности действий таковы (используем нумерацию действий в соответствии с п. 2.2):

«2323» и «3232» – эти цепочки характеризуют шаблон пищевого поведения,
 «1616» и «6161» – эти цепочки характеризуют шаблон оборонительного поведения,
 «2362», «3623» и «6236» – эти цепочки характеризуют шаблон «623», т. е. представляют собой смесь из первых двух шаблонов, фактически агент в этом случае «не определился» с ведущей потребностью и колеблется между двумя поведением, данные цепочки действий можно рассматривать как поисковое поведение,
 «4562» – эта цепочка характеризует шаблон поведения, направленного на размножение.

Таким образом, агент формирует цепочки действий, соответствующие удовлетворению его потребностей.

Была получена таблица средних весов правил для популяции из 100 агентов после 100 тыс. тактов времени (веса округлены до целых чисел). Таблица представлена на рис. 2. Таблица показывает, какое действие произведёт агент в каждой из возможных ситуаций: слева показаны три характеристики ситуации, а справа – веса каждого из возможных действий в этой ситуации. Видно, что выделяются некоторые правила, веса которых много больше остальных для такой же ситуации (характерные веса выделены жирным шрифтом). При ведущей пищевой мотивации преобладают действия, направленные на питание, при ведущей оборонительной – на оборону, при ведущей мотивации к размножению – действия, направленные на размножение. Также видны правила, образующие цепочки действий (ячейки весов выделены цветом): например,

при ведущей пищевой мотивации и отсутствии опасности преобладают правила: если предыдущее действие было «поиск пищи», то выполнить действие «питание», если предыдущее действие было «питание», то выполнить действие «поиск пищи».

Ситуация			Веса правил					
Ведущая мотивация	Предыдущее действие	Есть ли опасность	Покой	Поиск пищи	Питание	Подготовка к Размножению	Размножение	Оборона
Пищевая	Покой	Нет	18	4	18	73	14	19
Пищевая	Размножение	Нет	18	16	59	18	17	19
Пищевая	Поиск пищи	Нет	15	7	65	18	25	24
Пищевая	Оборона	Нет	16	68	10	12	19	19
Пищевая	Питание	Нет	20	73	18	6	1	9
Пищевая	Поиск пищи	Есть	16	16	54	16	14	18
Пищевая	Подготовка к размн.	Есть	14	3	13	11	13	24
Пищевая	Оборона	Есть	6	69	15	16	16	14
Пищевая	Покой	Есть	5	50	12	13	12	16
Пищевая	Подготовка к размн.	Нет	22	28	89	6	3	28
Пищевая	Питание	Есть	13	11	11	13	4	50
Пищевая	Размножение	Есть	13	3	12	6	13	22
Оборонительная	Питание	Есть	6	127	12	10	11	14
Оборонительная	Питание	Нет	28	21	17	19	13	121
Оборонительная	Подготовка к размн.	Нет	17	68	15	14	16	13
Оборонительная	Покой	Нет	25	30	27	26	28	54
Оборонительная	Оборона	Есть	15	71	1	16	14	9
Оборонительная	Поиск пищи	Есть	13	12	3	15	16	121
Оборонительная	Оборона	Нет	64	4	1	14	13	10
Оборонительная	Покой	Есть	6	16	13	14	15	12
Оборонительная	Подготовка к размн.	Есть	9	26	26	22	27	45
Оборонительная	Размножение	Нет	11	16	73	19	14	11
Оборонительная	Поиск пищи	Нет	16	19	6	13	18	120
Оборонительная	Размножение	Есть	20	19	13	17	22	43
К размножению	Оборона	Нет	12	15	13	242	3	12
К размножению	Питание	Есть	14	13	13	13	15	10
К размножению	Оборона	Есть	9	40	9	10	9	8
К размножению	Питание	Нет	20	21	7	17	7	10
К размножению	Подготовка к размн.	Нет	15	15	16	18	344	15
К размножению	Размножение	Есть	14	7	13	14	16	10
К размножению	Покой	Есть	13	12	10	16	13	15
К размножению	Поиск пищи	Нет	12	4	20	2	60	9
К размножению	Размножение	Нет	28	31	39	7	21	52
К размножению	Покой	Нет	15	14	18	18	14	14
К размножению	Поиск пищи	Есть	14	46	4	3	14	15
К размножению	Подготовка к размн.	Есть	11	20	30	36	22	26

Рис. 2. Веса правил системы управления, усреднённые по 100 агентам и округлённые до целых чисел.

Типичная динамика факторов F_F , F_S и F_R , соответствующих потребностям агента, приведена на рис. 3.

Видно, что при достаточно большом времени происходит максимизация факторов; все три фактора достигают порога, тем самым, все потребности агента удовлетворяются. Кроме того, наблюдается цикличность поведения агента, период цикла равен числу тактов времени между моментами размножения (период цикла приблизительно равен 40 тактам, характерный цикл показан участком «а» на рис. 3). В таком цикле агент сначала накапливает внутренний ресурс (осуществляется пищевое поведение), затем его действия направлены на максимизацию безопасности (оборонительное поведение) и поддержание уровня внутреннего ресурса, а когда обе потребности (пищевая и потребность безопасности) удовлетворены и прошло

некоторое время после предыдущего размножения, агент размножается (что соответствует потребности размножения).

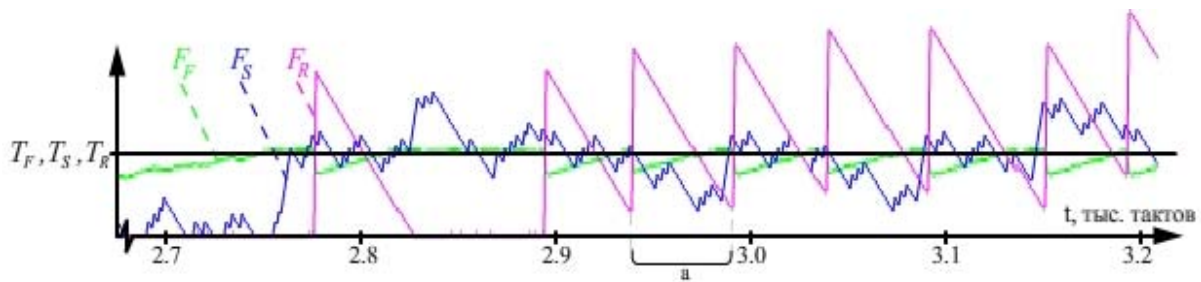


Рис. 3. Зависимость факторов F_F , F_S и F_R , соответствующих потребностям агента, от времени t . Участок «а» показывает типичный пример периода поведения агента. Отмечен уровень порогов T_F , T_S , T_R факторов.

На рис. 4 показана динамика ведущих мотиваций для того же агента.

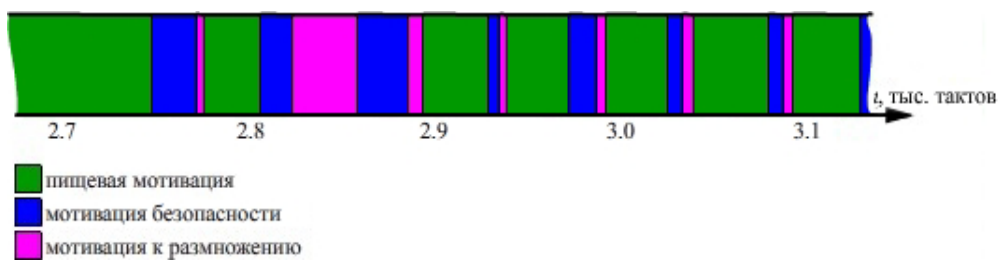


Рис. 4. Динамика ведущих мотиваций. На оси абсцисс указано время, цветом обозначена ведущая в данный момент времени мотивация.

Приведенные результаты показывают, что введение мотиваций так, как изложено в п. 2.3, обеспечивает естественное поведение агента: в системе управления агента определяется ведущая потребность, агент стремится её удовлетворить, демонстрируя соответствующий шаблон поведения. Если в процессе поведения выявилась более важная потребность, то агент переключается на неё. По достижении цели, когда фактор ведущей потребности достиг порога, выбирается следующая по значимости потребность. Использование факторов и порогов формирует чёткие последовательности действий, характер которых направлен на удовлетворение ведущей потребности.

3. ОБСУЖДЕНИЕ И ЗАКЛЮЧЕНИЕ

Таким образом, путем компьютерного моделирования исследовано поведение автономных агентов, обладающих несколькими потребностями и соответствующими потребностям мотивациями. Показано, что происходит формирование цепочек действий агентов. Продемонстрировано формирование циклов поведения, в которых последовательно удовлетворяются потребности питания, безопасности и размножения.

Интересно, что в настоящее время проводятся идейно близкие работы по моделированию поведения автономных агентов с несколькими потребностями и мотивациями [9]. Хотя в [9] не наблюдалось формирования цепочек действий и циклов поведения, как это представлено выше, в этой работе введена потребность поискового поведения, что позволяет моделировать отложенное обучение (первоначальное исследование лабиринта, а затем использование накопленного знания при поиске пищи), известное для живых организмов [10]. Отметим, что достаточно формальный подход к моделированию отложенного обучения был предложен и продемонстрирован

на простом примере также в работе [11]. Изложенная модель и работы [9,11] показывают, что возможно интересное направление перспективных теоретико-экспериментальных исследований с использованием двух взаимодополнительных подходов по изучению адаптивного поведения животных: 1) моделирования этого поведения и 2) биологических экспериментов по изучению поведения.

Кроме этого, отметим, что проведенные исследования могут рассматриваться как определенный вклад в перспективные работы по моделированию когнитивной эволюции [12].

СПИСОК ЛИТЕРАТУРЫ

1. Vernon D., Metta G., Sandini G. A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation (special issue on Autonomous Mental Development)*. 2007. V. 11. №. 2. P. 151–180.
2. Langley P., Laird J.E., Rogers S. Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*. 2009. V. 10. №. 2. P. 141–160.
3. Редько В.Г., Бесхлебнова Г.А. Моделирование адаптивного поведения автономных агентов. *Нейрокомпьютеры: разработка, применение*. 2010. № 3. С. 33–38.
4. Sutton R., Barto A. *Reinforcement Learning: An Introduction*. Cambridge: MIT Press, 1998.
5. Редько В.Г., Бесхлебнова Г.А. Модель формирования адаптивного поведения автономных агентов. В: *Интегрированные модели и мягкие вычисления в искусственном интеллекте: сборник научных трудов V-й Международной научно-практической конференции*. В 2-х томах. М.: Физматлит, 2009. Т. 1. С. 70–79.
6. Бесхлебнова Г.А., Редько В.Г. Модель формирования обобщенных понятий автономными агентами. В: *Четвертая международная конференция по когнитивной науке: тезисы докладов*. В 2-х томах. Томск: ТГУ, 2010. Т. 1. С. 174–175.
7. Непомнящих В.А., Попов Е.Е., Редько В.Г. Бионическая модель адаптивного поискового поведения. *Известия РАН. Теория и системы управления*. 2008. № 1. С. 85–93.
8. Редько В.Г., Непомнящих В.А. Бионическая модель поискового поведения. В: *Математическая биология и биоинформатика: Доклады II Международной конференции (7–13 сентября 2008 г., Пущино)*. Под ред. В.Д. Лахно. М.: МАКС Пресс, 2008. С. 128–129.
9. Butz M.V., Shirinov E., Reif K. Self-organizing sensorimotor maps plus internal motivations yield animal-like behavior. *Adaptive Behavior*. 2010. V. 18. №. 3–4. P. 315–337.
10. Tolman E.C. *Purposive behavior in animals and men*. New York: Appleton, 1932.
11. Witkowski M. An action-selection calculus. *Adaptive Behavior*. 2007. V. 15. №. 1. P. 73–97.
12. Редько В.Г. Моделирование когнитивной эволюции – перспективное направление исследований на стыке биологии и математики. *Математическая биология и биоинформатика*. 2010. Т. 5. № 2. С. 215–229. URL: [http://www.matbio.org/downloads/Redko2010\(5_215\).pdf](http://www.matbio.org/downloads/Redko2010(5_215).pdf) (дата обращения: 20.02.2012).

Материал поступил в редакцию 20.02.2012, опубликован 19.04.2012.