

Исследование механизмов целенаправленного адаптивного управления

М. С. Бурцев, Р. В. Гусарев, В. Г. Редько
Москва, ИПМ им. М. В. Келдыша РАН

Исследуется возможный механизм эволюционного возникновения целенаправленного адаптивного управления. Разработана и исследована компьютерная модель эволюции популяции агентов, обладающих двумя естественными потребностями: энергии и размножения. Каждая из потребностей характеризуется количественно мотивацией. Мотивации задают цели агентов, цели направлены на удовлетворение потребностей. Каждый агент имеет простую нейронную сеть, которая управляет его поведением. Модель демонстрирует, что целенаправленное поведение действительно возникает в ходе моделируемого эволюционного процесса. Возникновение целенаправленности сопровождается формированием простейшей иерархии в системе управления агентом.

Введение

Четкое понимание механизмов, обеспечивающих целенаправленность, необходимо при построении интеллектуальных систем управления сложными объектами. В представленной работе мы пытаемся выделить и исследовать некоторые общие функциональные свойства, определяющие целенаправленное адаптивное управление.

Для изучения понятия целенаправленности мы используем эволюционное компьютерное моделирование достаточно простой многоагентной системы. Подобное эволюционное моделирование многоагентных систем активно применяется в последнее время при исследовании экономических, информационных и социальных структур [1-3]. Мы также используем методы и концепции, которые развиваются в сравнительно новом направлении исследований "Искусственная жизнь" [3-5].

Основное понятие, которое мы используем для характеристики целенаправленного поведения - понятие мотивации. Под мотивацией подразумевается стремление к цели. Мы вводим количественные параметры, характеризующие мотивации агентов, и анализируем, как мотивации влияют на управление адаптивным поведением агентов и на эволюцию популяции агентов.

Основные предположения модели состоят в следующем:

- Имеется популяция агентов (искусственных организмов), имеющих естественные потребности (энергия, размножение).
- Популяция эволюционирует в простой клеточной среде. В клетках могут эпизодически появляться источники энергии (пища агентов). Каждый агент имеет внутренний энергетический ресурс, который пополняется при съедании пищи и уменьшается при выполнении агентом каких-либо действий. Уменьшение ресурса до нуля приводит к смерти агента. Агенты могут скрещиваться. Скрещивание приводит к рождению новых агентов.
- Каждая потребность характеризуется количественно мотивацией. Например, если энергетический ресурс агента мал, то есть мотивация найти пищу и пополнить энергетический ресурс.
- Поведение агента управляется его нейронной сетью, которая имеет специальные входы от мотиваций. Если имеется определенная мотивация, то поведение агента меняется с тем, чтобы удовлетворить соответствующую потребность. Такое поведение будем называть целенаправленным (есть цель удовлетворить определенную потребность).

1. Описание модели

1.1. Общая характеристика модели

Для простоты мы предполагаем, что среда, в которой живут агенты, одномерна. Агенты могут перемещаться в соседние клетки или перескакивать через несколько клеток.

Время дискретно. В каждый такт времени каждый агент популяции выполняет строго одно действие. Эти действия таковы:

- 1) находиться в состоянии покоя ("отдыхать");
- 2) двигаться, т.е. перемещаться на одну клетку вправо или влево;
- 3) прыгать через несколько клеток в случайно выбранном направлении;
- 4) есть (питаться);
- 5) скрещиваться.

Управление агентом (выбор действий) осуществляется его нейронной сетью. Параметры нейросети (веса синапсов) модифицируются в процессе эволюции популяции агентов. Веса синапсов нейронной сети составляют геном агента.

Эпизодически (случайно) на несколько тактов в некоторых клетках появляется пища агентов.

Агенты "близорукие", а именно, любой агент видит ситуацию только в трех клетках: в той, которой находится сам агент, и в двух соседних клетках. Мы определяем эти 3 клетки как "поле зрения" агента (рис. 1).



Рис. 1. Агенты в одномерной клеточной среде.

1.2. Взаимодействие агентов с окружающей средой и друг другом

Питание

Агент видит пищу, находящуюся в его поле зрения. Агент может съесть пищу в той клетке, в которой он находится. Если агент выработал действие "есть", и в его клетке есть пища, то внутренний энергетический ресурс агента увеличивается, а пища из клетки исчезает. Т.е. агент съедает всю пищу в клетке за один такт времени.

Скрещивание

Если два агента, находящиеся в соседних клетках, вырабатывают действие "скрещиваться", то возможно появление нового агента - "потомка". Геном (набор весов нейронной сети) потомка формируется на основе генома "родителей" при помощи рекомбинации и мутаций. Оба родителя передают часть своего энергетического ресурса потомку, полученная от родителей энергия составляет начальный ресурс потомка. Если энергетический ресурс одного или обоих родителей меньше ресурса, который необходимо передать потомку, появление нового агента не происходит, но ресурс родителей уменьшается, как и при появлении потомка, что приводит к гибели одного или обоих скрещивающихся агентов. После удачного скрещивания рожденный агент помещается в одну из клеток, соседних с клетками родителей, при условии, что хотя бы одна из этих клеток свободна (в ней нет другого агента). Если обе возможные клетки заняты, то потомок погибает. Последнее означает, что мир слишком тесен - нет "жизненного пространства" для потомка.

1.3. "Физиология" агентов

Опишем физиологию агентов, характеризуя взаимосвязи между действиями агента и его внутренним состоянием. Используем следующие обозначения: R - энергетический ресурс, ΔR - изменение ресурса за один такт времени; k_i - параметры, характеризующие изменение внутреннего энергетического ресурса; M_E , M_R - параметры, характеризующие мотивации.

Выполнение какого-либо действия приводит к изменению внутреннего ресурса агента по следующим правилам:

- 1) Отдых (нахождение в состоянии покоя): $\Delta R = -k_1$.
- 2) Движение (к соседней клетке): $\Delta R = -k_2$.
- 3) Прыжок (через несколько клеток): $\Delta R = -k_3$.
- 4) Питание: $\Delta R = -k_4 + k_5 A$, где A - количество пищи в клетке агента в данный такт времени.
- 5) Действие скрещивания: $\Delta R = -k_6$.

При появлении потомка родители передают ему часть своей энергии: $\Delta R_A = -k_7$, $\Delta R_B = -k_7$, где A и B - индексы родителей. Начальный ресурс потомка R_S равен полученной от родителей энергии $R_S = 2k_7$.

Прокомментируем приведенные соотношения.

Во время отдыха агент теряет небольшое количество ресурса (k_1). Двигаясь и прыгая, агент также затрачивает некоторое количество энергетического ресурса (k_2 , k_3). При питании агент теряет малое количество ресурса (k_4) и получает энергию, пропорциональную количеству пищи ($k_5 A$). Выполняя действия скрещивания, агент теряет в любом случае некоторую энергию (k_6). Если действия скрещивания агентов в соседних клетках скоординированы (оба скрещиваются), то рождается потомок, и оба родителя передают часть своей энергии (k_7) потомку.

Если энергетический ресурс агента недостаточен для выполнения выбранного действия, то этот агент умирает. Мотивации агента определяются следующим образом (см. рис. 2):

- 1) Мотивация к поиску пищи равна:

$$M_E = \max \left\{ \frac{R_0 - R}{R_0}, 0 \right\},$$

где R_0 - некое "оптимальное" значение энергетического ресурса.

2) Мотивация к скрещиванию составляет:

$$M_R = \min \left\{ \frac{R}{R_1}, 1 \right\},$$

где R_1 - "оптимальное" для размножения значение энергетического ресурса.

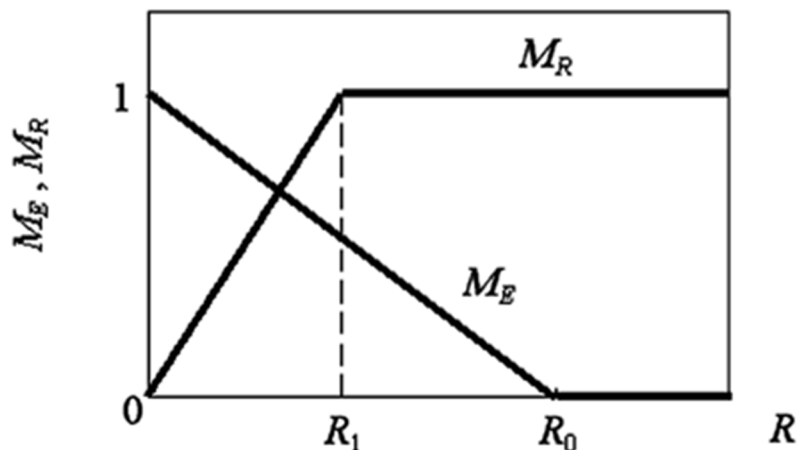


Рис. 2. Зависимость уровня мотиваций от внутреннего ресурса агента.

1.4. Структура нейронной сети агента

Нейронная сеть управляет поведением агента. Мы предполагаем, что нейронная сеть агента содержит один слой нейронов (рис. 3). На входы нейронной сети поступают сенсорные сигналы. Входы и нейроны связаны по принципу "все со всеми", то есть каждый нейрон получает сигналы от всех входов. Нейроны формируют выходные сигналы, которые определяют действия агента.

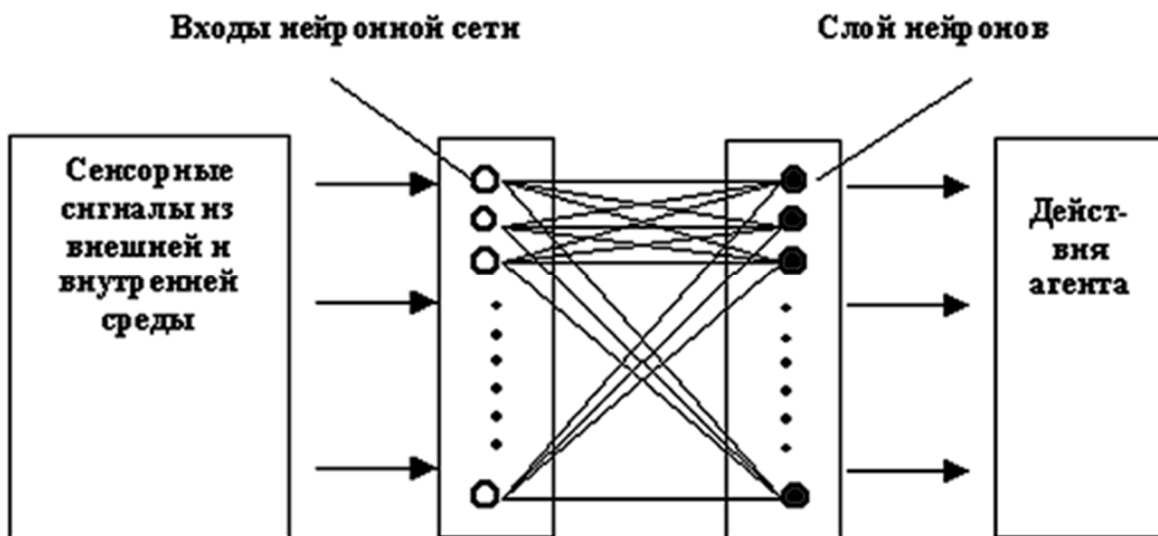


Рис. 3. Структура нейронной сети агента.

Рассмотрим входные сигналы нейронной сети.

Предполагаем, что агент получает из внешней среды следующую информацию:

1) наличие пищи во всех трех клетках поля зрения (рис. 1);

- 2) имеются ли агенты в клетках слева и справа;
- 3) мотивацию к скрещиванию соседних агентов в клетках слева и справа (считаем, что мотивация к скрещиванию определяет окраску агента [6], которую видят его соседи).

Кроме того, из внутренней среды агент получает информацию о своих мотивациях M_R и M_E .

Таким образом, мы имеем $3+2+2+2 = 9$ входных сигналов нейронной сети.

Нейроны определяют действия агента. Каждый нейрон соответствует одному действию. Будем считать, что агент в данный такт времени выполняет то действие, которое соответствует нейрону с максимальным выходным сигналом.

Действия агента перечислены в п. 1.1. Принимая во внимание, что действия движения и скрещивания имеют два варианта (двигаться и скрещиваться можно как направо, так и налево), получаем, что имеется 7 различных действий агента. Соответственно, нейронная сеть агента содержит 7 нейронов.

Мы предполагаем, что нейроны имеют логистическую активационную функцию:

$$y_j = F(\sum_i w_{ij} x_i), \quad F(b) = [1 + \exp(-b)]^{-1},$$

где y_j - выход j -го нейрона, x_i - входы нейрона, w_{ij} - синаптические веса j -го нейрона.

Так как входы и нейроны имеют полные межслойные синаптические связи (рис. 3), то нейронная сеть имеет $7 \times 9 = 63$ синаптических веса, которые полностью определяют работу системы управления агентом.

1.5. Схема эволюции

Популяция агентов эволюционирует во времени. Основной механизм эволюционного процесса - формирование генома рождающихся агентов. Геном кодирует веса связей нейронной сети агента, каждый вес представлен отдельным геном. Всего в нейронной сети имеется 63 веса, соответственно геном можно записать как:

$$S = (S_1, S_2, \dots, S_N), \quad N = 63.$$

При появлении нового агента, его геном формируется следующим образом:

- 1) осуществляется равномерная рекомбинация генов родителей;
- 2) получившийся геном подвергается малым мутациям.

Равномерная рекомбинация заключается в формировании каждого гена потомка путем случайного выбора этого гена от одного из двух его родителей. В процессе мутаций к каждому гену добавляется случайная величина z , равномерно распределенная в интервале $[-p_m, p_m]$:

$$S_i \rightarrow S_i + z_i, \quad i = 1, 2, \dots, N.$$

2. Моделирование

2.1. Параметры компьютерного эксперимента

Для того чтобы исследовать влияние мотиваций на поведение отдельного агента и популяции в целом, были проведены две серии экспериментов. В первой серии моделировалась эволюция популяции агентов с "работающими" мотивациями (введенными так, как это описано выше), во второй серии мотивации у агентов были "выключены" (входы от мотиваций искусственно "задавливались" специальным выбором параметров R_0 и R_1). В обоих случаях эксперимент проводился для нескольких значений вероятности случайного появления пищи в клетках, с тем, чтобы проанализировать влияние количества ресурса в окружающей среде на поведение популяции.

Из разумных соображений была задана физиология агентов, т.е. были зафиксированы параметры k_i , определяющие изменения энергетического ресурса R агента при совершении им различных действий. Эти параметры, оставшиеся неизменными во всех экспериментах, были заданы, исходя из следующих соображений:

- 1) при выполнении действия "отдыхать" (параметр k_1) агент тратит наименьшее количество запасенной энергии;
- 2) при выполнении действия "есть" (k_4) потери энергии в два раза больше, чем при выполнении действия "отдыхать";
- 3) при выполнении действия "двигаться" (k_2) потери энергии еще в два раза больше;
- 4) затраты на "скрещивание" (k_6) равны затратам на "прыжок" (k_3) и в 5 раз больше чем при "движении" (при "прыжке" агент перемещался на 5 клеток в случайную сторону).

Таким образом, соотношения для параметров, определяющих изменение внутреннего ресурса агента при выполнении действий, имеют следующий вид:

$$k_4 = 2k_1, k_2 = 2k_4, k_6 = k_3 = 5k_2.$$

Оптимальный ресурс агента R_0 задавался на три-четыре порядка больше, чем затраты на выполнение какого-либо действия ($R_0 = 10^4 k_1$). Таким образом, наличие оптимального ресурса позволяло агенту существовать несколько тысяч тактов времени. Значение параметра R_1 , определяющее мотивацию к скрещиванию, определялось, как $R_1 = 0,5R_0$.

Количество энергии, передаваемое от одного родителя к потомку, полагалось равным $k_7 = 0,1R_0$. Учитывая то, что потомок получает энергию от обоих, энергетический ресурс "новорожденных" агентов составлял $0,2R_0$.

Количество энергии, получаемое агентом в результате съедания "одной порции" пищи ($k_5 A$), было значительно больше типичной затраты на действия, но существенно меньше оптимального ресурса и составляло $0,02R_0$.

Пища сохранялась в клетке (если ее не съедал агент) в течение 20 тактов.

Интенсивность мутаций составляла $p_m = 0,05$.

Начальная популяция агентов состояла из 200 агентов.

Энергетический ресурс агентов начальной популяции задавался примерно в два раза больше оптимального значения R_0 .

Все агенты начальной популяции имели один и тот же набор синаптических весов.

Каждому синаптическому весу нейронной сети агента в нашей модели соответствует свой номер (от 1 до 63). Отличные от нуля веса нейронной сети агентов начальной популяции представлены в таблице 1. Эти веса были заданы, исходя из изложенных ниже соображений.

Мы требовали, чтобы агенты начальной популяции имели два основных "врожденных инстинкта": 1) инстинкт пополнения ресурса и 2) инстинкт размножения.

Инстинкт пополнения ресурса задавался весами 5, 13, 18 (таблица 1). Этот инстинкт определяет действия агента при наличии пищи в поле его зрения. Если агент видит пищу в соседней клетке, то он двигается к ней, если пища находится в одной с ним клетке, агент съедает ее.

При появлении в поле зрения другого агента включается инстинкт размножения, определяемый весами 23, 29. Этот инстинкт заключается в том, что наличие соседа вызывает действие скрещивания.

В дополнение к двум основным инстинктам был введен инстинкт "боязни тесноты", определяемый весами 24, 31. Этот инстинкт позволяет агентам сохранять свободу передвижения: если агент попадает в ситуацию, когда он с обеих сторон окружен соседями, то у него вырабатывается действие "прыгать".

Значения остальных синаптических весов (не указанных в таблице 1) для агентов начальной популяции полагались равными нулю.

Таблица 1. Отличные от нуля веса синапсов агентов начальной популяции.

Номер веса	Вход	Действие	Значение веса
5	Еда слева	Двигаться влево	0,7
13	Еда рядом	Есть	0,8
18	Еда справа	Двигаться вправо	0,7
23	Есть сосед слева	Скрещиваться влево	0,6
24	Есть сосед слева	Прыгать	0,4
29	Есть сосед справа	Скрещиваться вправо	0,6
31	Есть сосед справа	Прыгать	0,4

2.2. Результаты моделирования

Основным количественным показателем, по которому мы определяли "качество жизни" агентов, была общая численность популяции. На рисунке 4 представлены зависимости общей численности популяции от времени для агентов с мотивациями (мотивации "работают") и без мотиваций (мотивации "задавлены") при различном количестве пищи в нашей клеточной среде. Количество пищи задавалось параметром P_g - вероятностью случайного её появления в каждый такт времени в каждой клетке.

При малом количестве пищи (рис. 4а) популяция вымирает, как в случае агентов с мотивациями, так и для агентов без мотиваций - агенты довольно быстро расходуют начальный энергетический ресурс и не находят какой-либо стратегии выживания.

При среднем количестве пищи (рис. 4б) популяция успешно развивается, если эволюция находит достаточно эффективную стратегию поведения агентов. Для популяции агентов с мотивациями такая стратегия находится, а для популяции без мотиваций - нет, и популяция постепенно вымирает.

При избытке пищи (рис. 4в) агенты с мотивациями находят более эффективную жизненную стратегию, чем агенты без мотиваций: конечное число агентов в эксперименте с мотивациями более чем в полтора раза превышает численность популяции в эксперименте без мотиваций.

Приведенные результаты демонстрируют, что популяция агентов, имеющих мотивации, лучше приспосабливается к окружающей среде, чем популяция агентов без мотиваций.

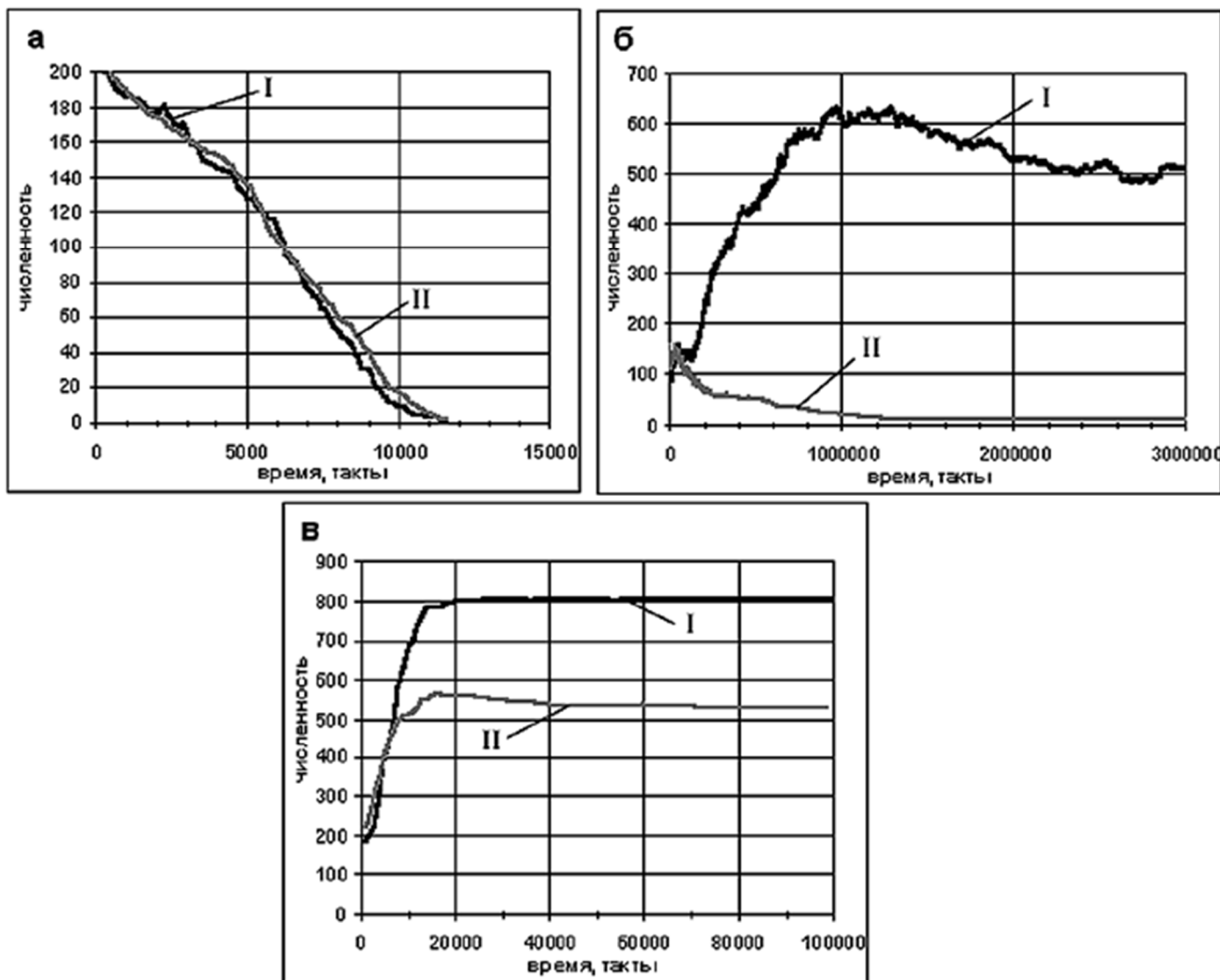


Рис. 4. Динамика численности популяции (начальная численность популяции 200 агентов, размер мира 900 клеток) для агентов с мотивациями (I) и без мотиваций (II) при различных вероятностях появления пищи P_g : а) $P_g = 1/2000$ (в любой клетке пища появляется случайно в среднем через 2000 тактов); б) $P_g = 1/200$; в) $P_g = 1/20$.

Рассмотрим, как формируется поведение агентов с мотивациями, для этого проанализируем зависимость от времени числа агентов, выполняющих отдельные действия на примере эксперимента со средним уровнем пищи (рис. 4б).

На рис. 5а,б показаны зависимости от времени численности агентов, выполняющих действия "отдыхать" и "питаться". Мы выбрали эти зависимости, так как большинство агентов в популяции выполняют именно эти действия. Рисунок 5в, на котором показана "нижняя часть" рис. 5б в увеличенном масштабе, дает представление об эволюции остальных действий агентов.

Видно, что на временном отрезке 0 - 10 тысяч тактов численность популяции быстро падает (с 200 примерно до 100), затем наблюдается небольшой рост. На отрезке 10 - 40 тысяч тактов доля "питающихся", составлявших подавляющее большинство в популяции в начале эволюции, сокращается, а доля "отдыхающих" растет. Это связано с тем, что действие "отдыхать" требует наименьших затрат энергии и, следовательно, когда агент не видит пищи или другого агента, ему выгодно выбирать именно это действие. На следующем этапе (50 - 150 тысяч тактов) в популяции агентов с мотивациями (рис. 5б) эволюция оптимизирует веса, связывающие нейроны агента с мотивационными входами, и затем начинается рост численности. В случае агентов без мотиваций такой оптимизации не происходит.

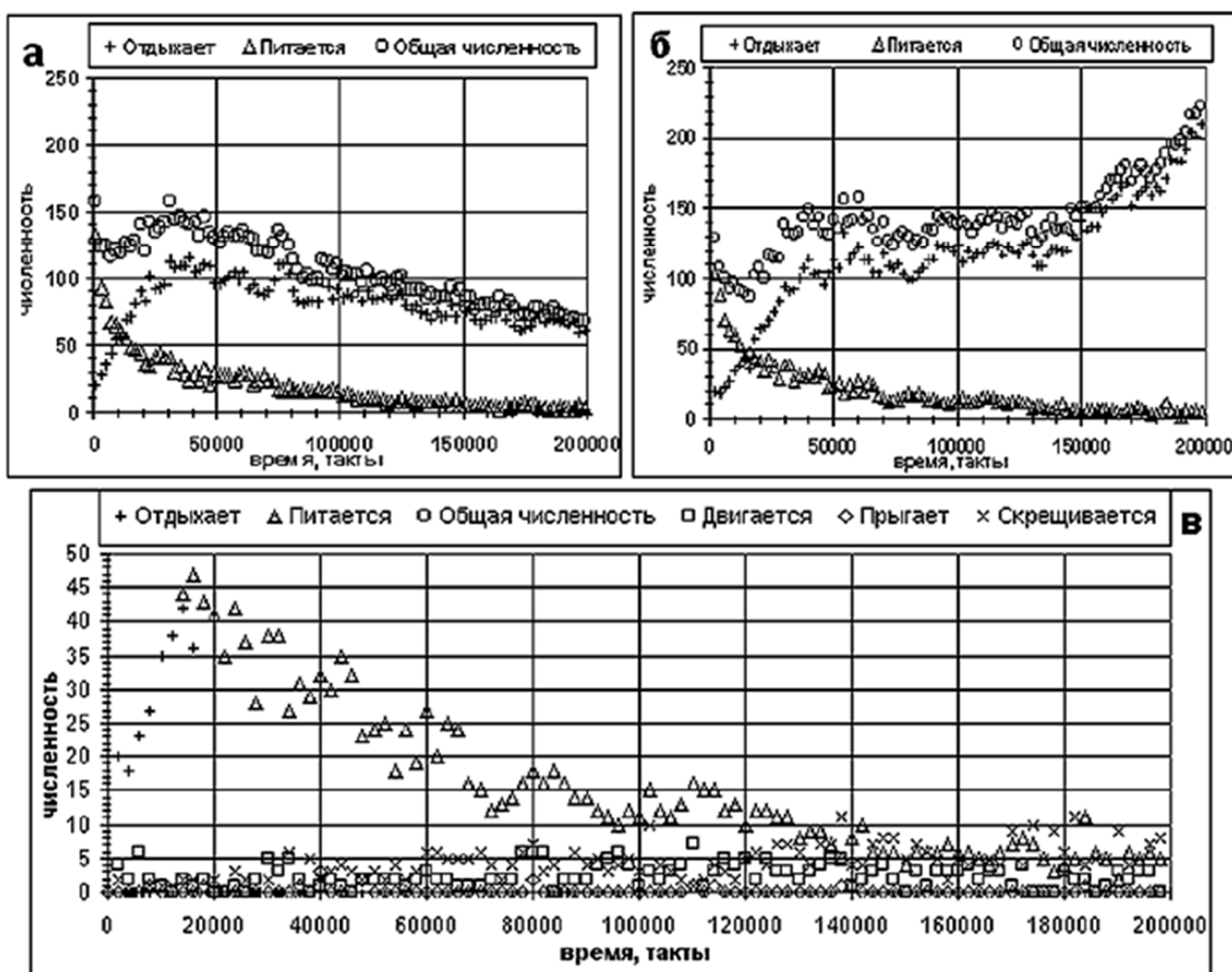


Рис. 5. Динамика распределения агентов по совершаемым действиям (эксперимент тот же, что и на рис. 4б): а) без мотиваций; б) с мотивациями; в) детали рис. 5б.

3. Анализ и обсуждение результатов

Результаты моделирования позволяют нам представить эволюцию системы управления поведением исследуемых агентов.

Рассмотрим сначала эволюцию стратегии поведения в популяции агентов, не обладающих мотивациями. Изначально у агентов имеются две заданные программы действий, направленные на пополнение ресурса:

1. "ВИЖУ ПИЦЦУ" -> "ЕСТЬ",

и на размножение:

2. "ВИЖУ СОСЕДА" -> "СКРЕЩИВАТЬСЯ",

а в ситуации, когда возбуждение входов нейросети агента не соответствует ни одному из инстинктов, выбор действия ничем не определен.

Через некоторое время эволюция вырабатывает для этого случая следующую программу:

3. "НЕ ВИЖУ НИЧЕГО" -> "ОТДЫХАТЬ".

Теперь действия определены для любых внешних условий, и стратегию поведения агента можно описать, как набор следующих простых рефлексов (рис. 6):

1. вижу пищу - есть;
2. вижу соседа, но не вижу пищу - скрещиваться;
3. не вижу ничего - отдыхать.



Рис. 6. Схема управления агента без мотиваций.

Видно, что действия агента напрямую определяются текущим состоянием окружающей среды.

Рассмотрим теперь эволюцию системы управления агента, обладающего мотивациями. Приведенные результаты, а также детальный анализ эволюции действий и весов нейронной сети агентов [7] позволяют представить следующую схему системы управления агентом. В дополнение к трем программам, рассмотренным выше, в процессе эволюции появляется следующий уровень управления - метапрограмма, управляющая выбором выработанных простейших программ в зависимости от мотиваций. Возникшая система управления является

иерархической - при помощи мотиваций формируется следующий уровень управления. Так в случае малого внутреннего ресурса мотивация к пополнению ресурса тормозит потребность в скрещивании, выключая программу 2. В случае большого внутреннего ресурса включается мотивация к размножению, что делает размножение предпочтительным, хотя и потребление пищи здесь также возможно. Возникшую при этом систему управления можно представить следующим образом (рис. 7):

1. мало ресурса - выполнять программы 1 и 3;
2. много ресурса - выполнять программы 1, 2 и 3.

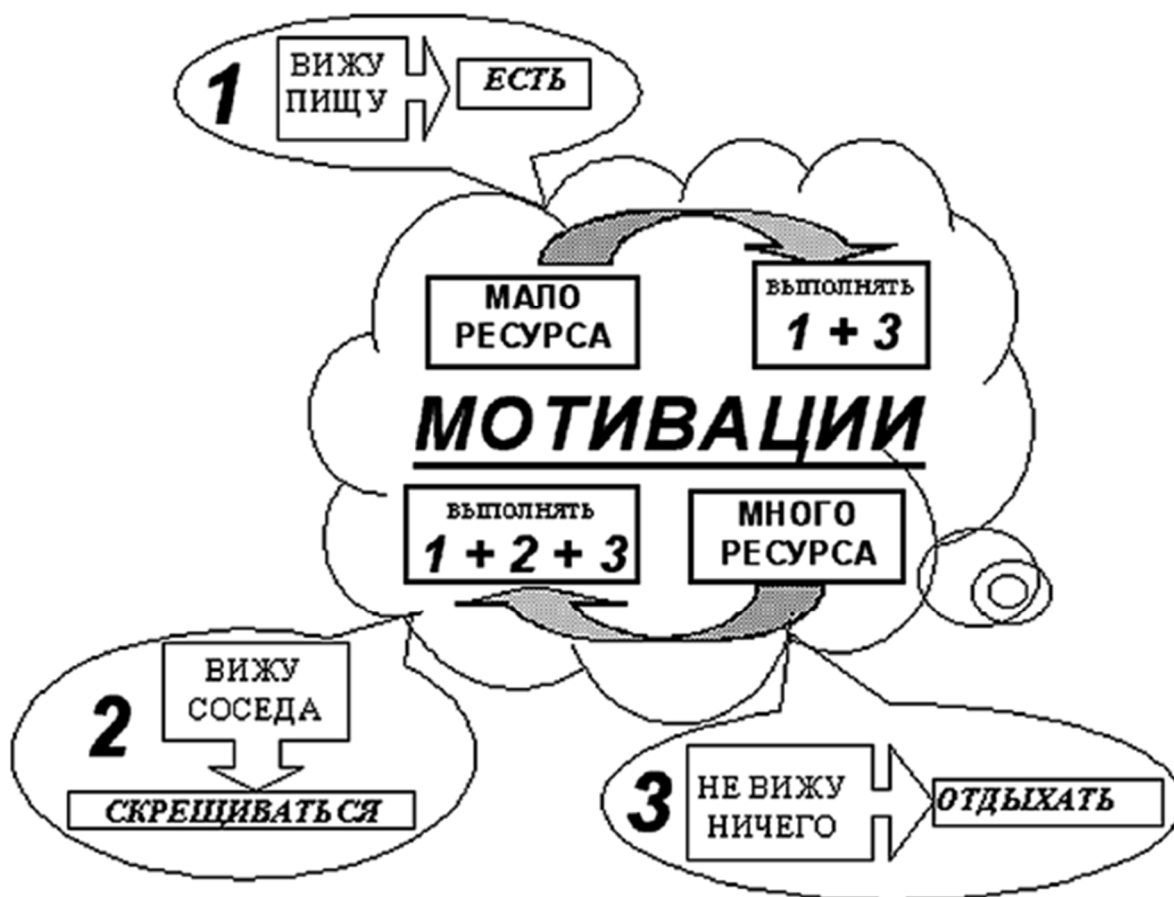


Рис. 7. Схема управления агента с мотивациями.

Возникновение нового уровня иерархии системы управления в нашей модели подобно метасистемному переходу от простых рефлексов к сложным рефлексам в теории метасистемных переходов В.Ф. Турчина [8].

Итак, представленные результаты демонстрируют, что поведение агента, управляемого иерархической системой, состоящей из двух уровней - уровня простых рефлексов и метауровня, обусловленного мотивациями, более эффективно, чем поведение, определяемое одними лишь простыми рефлексам. При этом мотивации формируют действия агента в соответствии с целями удовлетворения потребностей (в нашем случае потребностей питания и размножения).

Заключение

Хотя в представленной работе исследованы простейшие механизмы целенаправленного адаптивного управления для достаточно абстрактного примера, предложенный в работе подход

может служить основой для анализа целенаправленного управления в системах различной природы: робототехнических, производственных, экономических и социальных. В этом контексте отметим очень интересную и важную постановку широкого спектра исследований в рамках концепции "Эволюционной кибернетики", предложенной авторами (В.Ф. Турчин, Ф. Хейлигхен, К. Джослин) международного Интернет-проекта "Principia Cybernetica Project" [9]. Согласно этой концепции задача "Эволюционной кибернетики" [10] - "исследование возникновения и развития целенаправленных структур в Природе, Мышлении, Обществе и Технологиях".

Литература:

1. *Axelrod R.* The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration, Princeton: Princeton University Press, 1997
2. *Tesfatsion L.* "How Economists Can Get Alive" // W. Brian Arthur, Steven Durlauf, and David Lane (eds.), The Economy as an Evolving Complex System, II, Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, 1997, Volume XXVII.
См. также <http://www.econ.iastate.edu/tesfatsi/surveys.htm>
3. *Тарасов В.Б.* Искусственная жизнь и нечеткие эволюционные многоагентные системы - основные теоретические подходы к построению интеллектуальных организаций // Изв. РАН. Теория и системы управления. 1998. N.5.
4. *Langton C. G.* (Ed.) Artificial Life. // The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, Redwood City CA: Addison-Wesley, 1989.
5. *Редько В.Г.* Лекции по эволюционной кибернетике. Лекция 12. Направление исследований "Искусственная жизнь" - эволюционные и нейросетевые методы. // <http://www.keldysh.ru/pages/BioCyber/Lectures/Lecture12/Lecture12.html>
6. *Yaeger L.* Computational genetics, physiology, metabolism, neural systems, learning, vision, and behavior or Polyworld: Life in a new context. // Langton C.G. (ed). Artificial Life III. Addison-Wesley, 1994.
7. *Бурцев М.С., Гусарев Р.В., Редько В.Г.* Модель эволюционного возникновения целенаправленного адаптивного поведения 1. Случай двух потребностей. Препринт ИПМ РАН, 2000, N 43.
8. *Турчин В.Ф.* Феномен науки. Кибернетический подход к эволюции. М.: ЭТС, 2000.
9. *F. Heylighen, C. Joslyn, V. Turchin* (ed.) Principia Cybernetica Web (Principia Cybernetica, Brussels). // <http://pespmc1.vub.ac.be/>
10. *Heylighen F.* Evolutionary cybernetics // in: F. Heylighen, C. Joslyn and V. Turchin (editors): Principia Cybernetica Web (Principia Cybernetica, Brussels). <http://pespmc1.vub.ac.be/EVOLCYB.html> , 2000.