Бесхлебнова Г.А., Редько В.Г. Модели автономных адаптивных агентов // Сб. трудов НИИСИ РАН «Математическое и компьютерное моделирование систем: теоретические и практические аспекты», 2011 (в печати).

Модели автономных адаптивных агентов

Г.А. Бесхлебнова, к.т.н.
В.Г. Редько, д.ф.-м.н.

В работе сделан обзор наших компьютерных моделей адаптивного поведения автономных агентов. В первой модели исследованы автономные агенты, которые путем самообучения формируют свое поведение в двумерной клеточной среде. Показано, что агенты в процессе обучения способны самостоятельно формировать последовательные цепочки действий, а также понятия, обобщающие сенсорную информацию. Во второй модели продемонстрировано формирование достаточно естественного поведения агентов, обладающих потребностями питания, размножения, безопасности. Показано, что при эволюционной оптимизации систем управления агентов важную роль играет размножение.

The paper overviews authors' computer models of adaptive behavior of autonomous agents. The first model investigates autonomous self-learning agents that form their behavior in two-dimensional cellular environment. It is shown that the agents during learning are able to form the sequential chains of actions and the notions summarizing the sensory information. The second model demonstrates the formation of natural behavior of agents that have needs of feeding, reproduction, and safety. It is shown that reproduction plays an important role at evolutionary optimization of control systems of agents.

1. Введение

Одно из новых и интересных направлений, которое развивается в последние годы в вычислительном интеллекте (Computational Intelligence), – исследование и применение автономных адаптивных агентов [5, 9]. Такие агенты, подобно живым организмам, могут обладать собственными целями, собственными знаниями, формировать собственную политику поведения, выполнять те или иные действия, а также взаимодействовать с другими агентами. В связи с этим важно и интересно исследовать свойства автономных адаптивных агентов.

Также исследование автономных агентов может составлять начальный этап исследований когнитивной эволюции – эволюции познавательных способностей биологических организмов, приведшей к возникновению способностей научного познания [2].

В настоящей статье излагаются результаты выполненных в последние годы в ЦОНТ НИИСИ РАН исследований адаптивных свойств автономных агентов. Излагаются две модели автономных агентов: модель формирования адаптивного поведения автономных агентов в двумерной клеточной среде (раздел 2) и модель автономных агентов, обладающих естественными потребностями (раздел 3).

2. Модель формирования адаптивного поведения автономных агентов в двумерной клеточной среде

В данном разделе излагается модель поведения автономных агентов в двумерной клеточной среде (мире) [1,3]. Предполагается, что в любой клетке мира может находиться только один агент. У каждого агента задано свое направление «вперед». В некоторых клетках, число которых фиксировано, расположена пища агентов, величина порции пищи в каждой из этих клеток тоже фиксирована. Агент обладает ресурсом $R(t)$, где t – текущий момент времени. Ресурс агента увеличивается при питании агента, при выполнении агентом других действий ресурс уменьшается. Если ресурс агента $R(t)$ в результате его действий становится меньше определенного порога R_{min}, то данный агент умирает. Агенты функционируют в дискретном времени, $t = 0,1,\ldots$

Агенты обладают простыми когнитивными способностями: они запоминают закономерности взаимодействия с внешней средой в системе логических правил вида «Если имеет место ситуация $\mathbf{S}(t)$, то следует выполнить действие $A(t)$».

Каждый такт времени агент выполняет одно из следующих семи действий: деление, питание, перемещение на одну клетку вперед, поворот направо или налево, нанесение удара по агенту, находящемуся впереди данного, отдых. Система управления агента основана на классифицирующих системах [6], представляющих собой набор логических правил, формируемых как в процессе эволюции популяции, так и путем самообучения агентов. Этот набор правил составляет геном агента. Каждый такт времени агент осуществляет выбор действия и обучается.

Действие «деление» происходит следующим образом: в одной из соседних клеток, случайно выбираемой, рождается потомок агента; если все соседние клетки данного агента заняты, то потомок не рождается; геном рождающегося потомка отличается от генома родителя случайными мутациями. При делении агента ресурс родителя делится пополам между родителем и потомком. Логические правила потомка отличаются от правил родителя малыми мутациями.

При выполнении действия «питание» агент съедает всю порцию пищи в той клетке, в которой он находится. После этого новая порция пищи помещается в случайную клетку.

Если агент ударяет находящегося впереди него другого агента, то нападающий агент отнимает у ударяемого определенную величину ресурса. Если оба агента нападают друг на друга, то ресурс обоих уменьшается на величину, расходуемую на действие «ударить».

Размер двумерного мира равен N_xN_y клеток (координаты клеток равны $x = 1,\ldots,N_x$; $y = 1,\ldots,N_y$). Клеточный мир замкнут: если агент, находящийся в клетке с координатой $x = N_x$, движется вправо, т.е. пересекает «границу мира», то он перемещается в клетку с координатой $x = 1$, аналогично происходит движение агента при пересечении других границ мира.

Выбор действий агента обеспечивается имеющейся у него системой управления. Система управления агента представляет собой набор правил вида:

$$\mathbf{S}_k(t) \rightarrow A_k(t) , \qquad (1)$$

где $\mathbf{S}_k(t)$ – текущая ситуация, $A_k(t)$ – действие, соответствующее этому правилу, k – номер правила. Каждое правило имеет свой вес W_k , веса правил модифицируются при обучении агента. $\mathbf{S}_k(t)$ есть вектор, компоненты которого принимают значения 0 либо 1. Значения 0 и 1 соответствуют наличию или отсутствию порции пищи или другого агента в определенной клетке в «поле зрения» агента. Поле зрения включает в себя четыре клетки: ту клетку, в которой агент находится, клетку впереди агента и клетки справа и слева от агента.

Обозначим A^* намечаемое к выполнению действие. Это действие может быть выбрано либо в соответствии с имеющимися у агента правилами, либо случайным образом. Выбор действия агентом осуществляется следующим образом. Определяется текущая ситуация $\mathbf{S}(t)$ и формируется множество выделенных правил $\{\mathbf{R}_\mathbf{S}\}$, в это множество включаются те правила агента, для которых все компоненты вектора $\mathbf{S}_k(t)$ совпадают с компонентами вектора $\mathbf{S}(t)$, т.е. $\mathbf{S}_k(t) = \mathbf{S}(t)$. Из правил, входящих в $\{\mathbf{R}_\mathbf{S}\}$, выбирается правило, для которого вес правила W_k максимален, и с вероятностью $1\text{-}\varepsilon$ намечается для выполнения действие $A^* = A_k(t)$, входящее в это правило, а с вероятностью ε для выполнения намечается случайное действие A^*. Если правил, для которых $\mathbf{S}_k(t) = \mathbf{S}(t)$, у данного агента нет, т.е. множество $\{\mathbf{R}_\mathbf{S}\}$ оказывается пустым, то намечается для выполнения случайное действие A^*. Если при случайном выборе действия у рассматриваемого агента правило $\mathbf{S}(t) \rightarrow A^*$ отсутствует, то это новое правило добавляется к имеющимся, вес его полагается равным 0. В результате для выполнения намечается действие A^* и формируется новое правило (если у агента не было правила, соответствующего текущей ситуации $\mathbf{S}(t)$ и намеченному действию A^*). Далее намеченное действие A^* выполняется.

При моделировании часто использовался «метод отжига»: на начальных тактах моделирования, когда логические правила агентов еще не сформированы, полагалось $\varepsilon \sim 1$, т.е. в любом случае была большая вероятность случайного выбора действий, со временем величина ε по экспоненте уменьшалась до нуля, характерное время уменьшения ε составляло 1000 тактов времени. После этого выбор действия осуществлялся в соответствии с правилами и их весами.

При обучении веса правил W_k модифицировались методом обучения с подкреплением [8], т.е. путем самообучения агента на основе изменения его ресурса. Изменение весов W_k при обучении происходило следующим образом. Менялся вес того правила, которое использовал агент в предыдущий такт времени t-1, этот вес изменялся в соответствии с изменением ресурса агента при переходе к такту t и весом правила, применяемого в такт t. Пусть вес правила, примененного в такт t-1, равен $W(t\text{-}1)$, вес правила, применяемого в такт t, равен $W(t)$, ресурс агента в эти такты времени равен $R(t\text{-}1)$ и $R(t)$, соответственно. Тогда изменение веса $W(t\text{-}1)$ равно [8]:

$$\Delta W(t\text{-}1) = \alpha\,[R(t) - R(t\text{-}1) + \gamma W(t) - W(t\text{-}1)] , \qquad (2)$$

где α – параметр скорости обучения, γ – дисконтный фактор; $0 < \alpha << 1$, $0 < \gamma < 1$, $1\text{-}\gamma << 1$. В результате

обучения увеличивались веса правил, применение которых приводило к росту ресурса агента.

Моделирование проводилось в рамках полной описанной модели и в рамках упрощенной версии. В последнем случае изучалось обучение одного агента. Изложим результаты моделирования.

В случае полной версии модели рассматривалась популяция, состоящая из $n = 50$ агентов, помещенная в мир из 100 клеток ($N_x = N_y = 10$), в котором в половине клеток была случайно распределена пища. В этом случае агент определял наличие/отсутствие пищи в 4-х клетках поля зрения и наличие/отсутствие другого агента в 3-х клетках поля зрения (впереди, справа, слева), т.е. каждая ситуация $\mathbf{S}(t)$ характеризовалась бинарным вектором, имеющим семь компонент. Следовательно, всего было 128 возможных ситуаций и 7 возможных действий; итого, имелось 896 возможных правил. Было продемонстрировано, что в полной модели в процессе эволюции и обучения агентов формировалось естественное их поведение: агенты преимущественно питались и часто отнимали ресурс друг у друга, изредка они выполняли и другие действия.

Прирост ресурса агента при съедании пищи был равен 1 (считаем ресурс безразмерным). Расход ресурса на каждое из действий, кроме удара, был равен 0.01, расход на удар составлял 0.02, при ударе нападающий агент отнимал у ударяемого ресурс, равный 0.05. Параметры обучения с подкреплением были следующими: $\alpha = 0.1$, $\gamma = 0.9$. Применялся метод отжига. Интенсивность мутаций была равна 0. Исходный ресурс агента (при $t = 0$) составлял $R = 1$. Минимальный ресурса агента $R_{\min}$ (при $R < R_{\min}$ агент умирал) был равен 0.

На рис. 1 представлена зависимость среднего по популяции ресурса агента $\langle R \rangle$ от номера такта времени t. Видно, что сначала ($t < 2000$) скорость роста $\langle R \rangle$ мала, так как логические правила еще не сформированы, и веса имеющихся правил еще не настроены. При $t > 5000$ скорость роста $\langle R \rangle$ практически постоянна; стохастичность изменения $\langle R \rangle$ на этом участке обусловлена случайным перемещением агентов по клеткам мира, а также случайным размещением новых порций пищи в ячейках мира после выполнения агентами действия «питание».

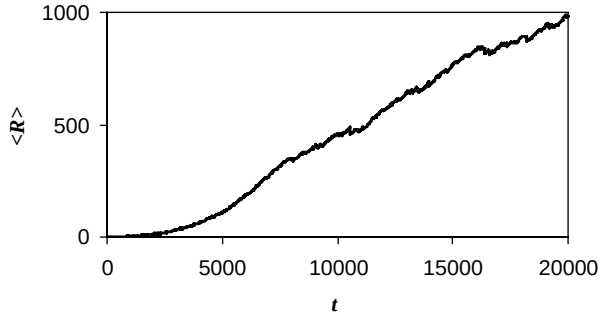

Рис. 1. Зависимость среднего по популяции ресурса агента $\langle R \rangle$ от номера такта времени t в случае полной модели

На рис. 2 представлена зависимость числа агентов Ne, выполняющих действие «питание», от номера такта времени t. Видно, что при больших значениях t примерно 30% агентов (из общего числа агентов популяции, равного 50) выполняет каждый такт это действие. Наблюдаются сильные стохастические колебания числа Ne во времени. Примерно такая же зависимость от времени наблюдается и для числа агентов, выполняющих действие «нанесение удара»; число таких агентов при больших t равно примерно 15-20. При больших t число агентов, выполняющих действие «деление», мало и составляет около 1, а число агентов, выполняющих каждое из остальных действий (движение вперед, повороты направо/налево, отдых), равно примерно 3-5.

Таким образом, в изложенной полной модели агенты обучились выполнять преимущественно действия «питание» и «нанесение удара», которые приводили к увеличению ресурса агента, и научились уклоняться от выполнения действия «деление», которое приводило к уменьшению ресурса (ресурс делящегося агента уменьшался в 2 раза). Каждое из остальных действий выполнялось с небольшой частотой. Это позволяет говорить о том, что в модели эволюции популяции самообучающихся агентов формируется естественное поведение.

В упрощенной версии модели в клеточном мире оставался один самообучающийся агент. Изучался вопрос: может ли агент формировать цепочки действий? Для обучения такого агента использовался метод обучения с подкреплением. Рассматривалось два варианта формирования цепочек. В первом варианте агент мог выполнять только 4 действия: питаться, двигаться вперед и поворачиваться направо либо налево. Считалось, что имелась только одна расположенная в определенной клетке порция пищи. Агенту необходимо было сформировать заданную цепочку из одного, двух или трех действий. Например, трехзвенная цепочка включала следующие действия: 1) «поворот направо», 2) «перемещение вперед», 3)

«питание»; при этом порция пищи исходно располагалась в клетке справа от той клетки, в которую исходно помещался агент. Основные параметры расчета были такими же, как и для полной модели. Метод отжига в этом варианте не использовался. Величина ε, регулирующая случайный выбор действия агентом, была постоянной и составляла $\varepsilon = 0.2$. Расчеты показали, что простые одно-, двух-, трехзвенные цепочки действий достаточно легко формировались в процессе самообучения агента.

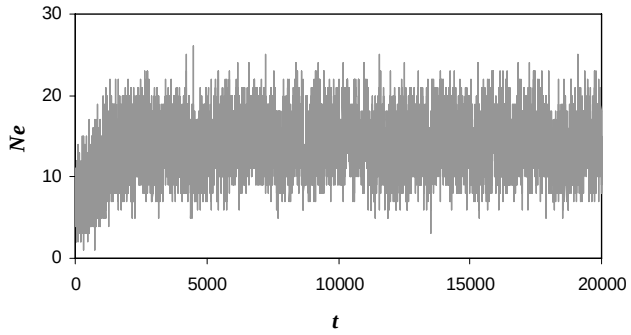

Рис. 2. Зависимость числа агентов *Ne*, выполняющих действие «питание», от номера такта времени *t* в случае полной модели

Во втором варианте упрощенной версии к указанным 4-м действиям добавлялось еще действие «отдых». Как и в полной модели, в половине клеток была случайно распределена пища. Применялся метод отжига (характерное время уменьшения ε составляло 1000 тактов времени). Основные параметры расчета были такими же, как в полной модели. Расчеты показали, что и в этом случае формировались заранее неизвестные цепочки действий из нескольких звеньев, приводящие к нахождению пищи. Пример зависимости ресурса *R* агента от времени для данного случая показан на рис. 3.

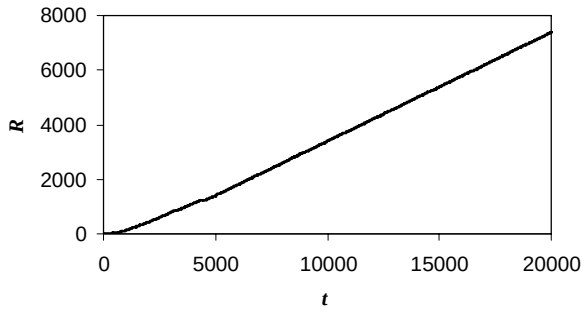

Рис. 3. Зависимость ресурса *R* отдельного самообучающегося агента от номера такта времени *t* в случае упрощенной версии модели

Поскольку агент был один, то каждая ситуация $\mathbf{S}(t)$ определялась только наличием/отсутствием пищи в 4-х клетках поля зрения и характеризовалась бинарным вектором, имеющим 4 компоненты. Всего было 16 возможных ситуаций и 5 возможных действий; итого, имелось 80 возможных правил. Интересно, во всех расчетах общее число правил, сформированных каждым агентом, было равно 80. В начале расчета, когда вероятность случайного выбора действия была высока, агент путем случайного поиска формировал все возможные правила. Однако веса этих правил менялись в процессе обучения, и к концу расчета преимущественно использовались только 16 правил из 80 возможных.

В конце расчета были выделены логические правила, имеющие достаточно большой вес (превышающий 1). Оказалось, что для типичного расчета число таких выделенных правил с большими весами было равно 16, каждое из правил соответствовало одной из возможных ситуаций и выполняемому в этой ситуации действию. Именно эти правила применялись агентом. Этот набор правил можно рассматривать как обобщающие эвристики, формируемые агентом в процессе самообучения. Эти эвристики сводятся к следующему: 1) если порция пищи расположена в той же клетке, в которой находится агент, то нужно выполнить действие «питание» (таких правил было 8); 2) если порции пищи нет в той клетке, в которой находится агент, и пища находится в клетке впереди или справа/слева от агента, то нужно выполнить

действие «перемещение вперед» или «поворот направо/налево», соответственно; если вообще не было пищи в поле зрения агента, то агент предпочитал действие «перемещение вперед». Перемещение вперед имело предпочтение перед поворотами, это можно объяснить тем, в первом случае образуется двухзвенная цепочка действий, а во втором – трехзвенная. Интересно, что когда агент попадал в ситуацию «буриданова осла», т.е. наблюдал пищу в двух клетках: справа и слева, то в одних расчетах он предпочитал поворачиваться направо, а в других – налево. Отметим, что действие «отдых» игнорировалось во всех ситуациях. В некоторых расчетах было небольшое число и других правил с весами, большими 1, тем не менее, свойства применяемых правил только изредка незначительно отличались от вышеописанных. Представленная на рис. 3 зависимость $R(t)$ соответствовала 16 выделенным правилам с большими весами: для каждой из возможных 16 ситуаций было свое правило, характер этих правил изложен выше. Следовательно, в процессе обучения агент самостоятельно формировал вполне естественные правила, определяющие «разумную» стратегию поведения.

Одно из важных когнитивных свойств живых организмов – свойство формирования обобщенных понятий. Использование понятий приводит к сокращению требуемой памяти и времени обработки информации. Но как происходит формирование понятий, можно ли представить процессы формирования понятий с помощью компьютерного моделирования? В изложенной выше модели продемонстрировано, что в процессе обучения агент выбирает определенные правила, использование которых можно рассматривать как поведение в соответствии с обобщающими эвристиками. Однако эти эвристики были обнаружены путем наблюдения за работой компьютерной программы, у самих же агентов эти эвристики в явном виде не формировались. Для того чтобы продемонстрировать способность агента самостоятельно проводить обобщения и формировать понятия, в компьютерную программу дополнительно вводилась процедура усреднения. Агент для каждого действия проводил усреднение по времени, а именно, вычислялось среднее число применений данного действия по всему времени расчета для той или иной текущей ситуации $\mathbf{S}(t)$. Это позволяло агенту производить обобщение ситуаций и формировать понятия, характеризующие внешнюю среду. Причем эти понятия важны для агента, так как они связаны с его действиями и с приростом его ресурса.

Проведенное усреднение показало, что оцененные в конце расчета средние по времени частоты действий определяют следующие преимущественные действия агента. Действие *питание* выполняется, если имеется пища в той клетке, в которой находится агент (независимо от того, имеется ли пища в других клетках поля зрения агента). Действие *перемещение на одну клетку вперед* выполняется, если нет пищи в той клетке, в которой находится агент, и имеется пища в клетке впереди агента. Действие *поворот направо/налево* выполняется, если нет пищи в той клетке, в которой находится агент, и в клетке впереди агента, но имеется пища в клетке справа/слева от агента. Частота действия *отдых* пренебрежимо мала. Таким образом, формировались цепочки действий, приводящие к нахождению пищи и увеличению ресурса агента.

Можно говорить о том, что агент формирует внутренние понятия «*имеется пища в моей клетке*», «*имеется пища в клетке впереди меня*», «*имеется пища в клетке справа/слева от меня*». Наблюдая за ситуациями и выполняемыми действиями, агент способен самостоятельно формировать понятия, обобщающие сенсорную информацию.

3. Модель автономных агентов, обладающих естественными потребностями

Современные исследования адаптивного поведения и интеллектуальных систем включают активные работы по изучению автономных адаптивных агентов и их интеллектуальных и когнитивных свойств (см., например, [9]). Тем не менее, фактически отсутствуют работы по моделированию процессов формирования адаптивного поведения агентов с естественными потребностями. В настоящем разделе излагается модель адаптивного поведения автономных агентов, обладающих потребностями, аналогичными основным потребностям биологических организмов: питание, размножение, безопасность [4].

В модели предполагается, что мир состоит из двух клеток: опасной и безопасной. С периодом T_D тактов времени статус клеток меняется: опасная клетка становится безопасной, и, наоборот, клетка, бывшая безопасной, становится опасной. Агент, находящийся в опасной клетке, каждый такт времени теряет ресурс r_D. Сенсорная система агента определяет ситуацию $\mathbf{S}(t)$, характеризующую внешнюю и внутреннюю среду агента. Вектор $\mathbf{S}(t)$ состоит из 3 компонент, принимающих значения 0 либо 1 и определяющих следующее: 1) имеется ли в мире достаточное количество пищи, превышающее определенный порог f_{th}, 2) превышает ли собственный ресурс агента $R(t)$ заданный порог r_{th}, 3) опасна ли клетка, в которой находится агент. Таким образом, имеется 8 различных ситуаций $\mathbf{S}(t)$.

Каждый такт времени агент выполняет одно из следующих 4-х действий: деление, питание, перемещение в другую (альтернативную из двух) клетку, отдых.

Выбор действий агента обеспечивается имеющейся у него системой управления. Система управления агента представляет собой набор правил вида (1). Веса правил W_k модифицируются при обучении агента

методом обучения с подкреплением. Так как общее число различных ситуаций равно 8, а число действий равно 4, то всего имеется 32 различных правила. Начальный набор весов этих правил $\{W_{0k}\}$, получаемый агентом от родителя (с небольшими мутациями), представляет собой геном агента. В противоположность геному текущие веса правил $\{W_k\}$, которые использует агент при выборе действия, модифицируются при обучении агента. Таким образом, каждый агент имеет два набора весов правил: начальные веса $\{W_{0k}\}$, составляющие геном агента и не меняющиеся в течение его жизни, и текущие используемые веса $\{W_k\}$, модифицируемые при жизни агента путем обучения. В момент рождения агента текущие веса равны начальным: $\{W_k\} = \{W_{0k}\}$.

Чтобы учесть ограничение на возраст агентов, считается, что агент с определенной вероятностью P_d ($P_d << 1$) может погибнуть каждый такт времени (от случайных факторов), это соответствует средней продолжительности жизни агентов порядка $1/P_d$ тактов времени.

Опишем действия агента. Действие «деление» происходит следующим образом: рождается потомок данного агента, ресурс родителя делится пополам между родителем и потомком; геном $\{W_{0k}\}$ рождающегося потомка отличается от генома родителя случайными мутациями.

При выполнении действия «питание» агент съедает определенную часть r_{eat} пищи, если такое количество пищи имеется в данный такт времени в мире. Ресурс агента увеличивается на величину r_{eat} .

При выполнении одного из действий «деление», «питание», «перемещение» и «отдых» ресурс агента уменьшается на величину r_d , r_e , r_t , r_r , соответственно. Действия «деление» и «питание» соответствуют потребностям размножения и питания. Действие «перемещение» соответствует потребности безопасности, так как оно может обеспечить движение агента из опасной клетки в безопасную; моделирование показывает, что такое обеспечение действительно происходит.

Каждый такт времени агент осуществляет выбор действия и обучается. При выборе действия агента определяется текущая ситуация $\mathbf{S}(t)$ и выделяются 4 правила, для которых $\mathbf{S}_k = \mathbf{S}(t)$. Далее используется ε-жадный метод: с вероятностью $1-\varepsilon$ из этих выделенных правил выбирается то, для которого вес W_k максимален, а с вероятностью ε – произвольное из этих правил ($1 >> \varepsilon > 0$). Действие A_k , соответствующее выбранному правилу, выполняется.

Основные параметры расчета были следующими. Максимальная численность популяции составляла $N_{max} = 100$ (если численность популяции достигала величины N_{max} , то новые агенты не рождались). Расход ресурса на каждое из действий (r_d , r_e , r_t , r_r) был равен 0.01. Период смены статуса клеток (опасная ↔ неопасная) составлял $T_D = 100$ тактов времени. Уменьшение ресурса агента за один такт времени при нахождении его в опасной клетке было равно $r_D = 10$. Увеличение ресурса агента при питании составляло $r_{eat} = 10$. Вероятность гибели агента от случайных факторов составляла $P_d = 0.001$. Параметры обучения с подкреплением: $\alpha = 0.1$, $\gamma = 0.9$. Параметр ε-жадного метода при случайном выборе правила составлял $\varepsilon = 0.1$. Изменение порогов f_{th}, r_{th} не сильно влияло на поведение агентов; в типичных расчетах эти величины составляли: порог значимого количества пищи в мире $f_{th} = 10$, порог значимого собственного ресурса агента $r_{th} = 1$. Система управления каждого агента состояла из всех 32-х возможных правил, в начале расчета веса правил W_{0k} , составляющие геном агента, были случайными и малыми по сравнению с последующими значениями W_{0k}. При мутациях к начальным весам правил рождающихся агентов W_{0k} добавлялась случайная величина, равномерно распределенная в интервале $[-0.5P_m, 0.5P_m]$; $P_m = 0.1$ – интенсивность мутаций. В каждый такт времени при выборе действия каждого агента с вероятностью 0.5 в мир добавлялась порция пищи, равная 10.

Специальным выбором параметров задавались следующие случаи:

- Случай L (чистое обучение); в этом случае интенсивность мутаций полагалась нулевой: $P_m = 0$.
- Случай E (чистая эволюция), в этом случае интенсивность обучения была нулевой, вероятность выбора случайного правила также обнулялась: $\alpha = 0$ и $\varepsilon = 0$.
- Случай LE (обучение + эволюция), т.е. полная модель, с приведенными выше параметрами.

На рис. 4 представлены зависимости среднего по популяции ресурса агента $\langle R \rangle$ от номера такта времени t для случаев чистого обучения L и чистой эволюции E. На рис. 5 приведена зависимость $\langle R \rangle(t)$ для случая E при больших временах. Зависимости усреднены по 100 различным расчетам, выполненным для разных используемых последовательностей случайных чисел.

Согласно рис. 4, 5 в случае эволюционной оптимизации при больших значениях t средний ресурс постепенно уменьшается, а в случае обучения $\langle R \rangle$ при больших t становится постоянным. Только при $t > 100000$ (случай E) наблюдается выход на асимптотическое значение $\langle R \rangle$, однако меньшее, чем асимптотическое значение для случая L. Анализ кривых $\langle R \rangle(t)$ при $t < 1000$ показывает, что в случае L рост величины $\langle R \rangle$ происходит примерно в два раза быстрее, чем в случае E.

В случае полной модели LE зависимость $\langle R \rangle(t)$ близка к таковой в случае чистого обучения L.

Таким образом, анализ зависимостей $\langle R \rangle(t)$ показывает, что за исключением времени локального увеличения величины $\langle R \rangle$ при $100000 > t > 1000$ чистое обучение имеет преимущество перед эволюционной оптимизацией.

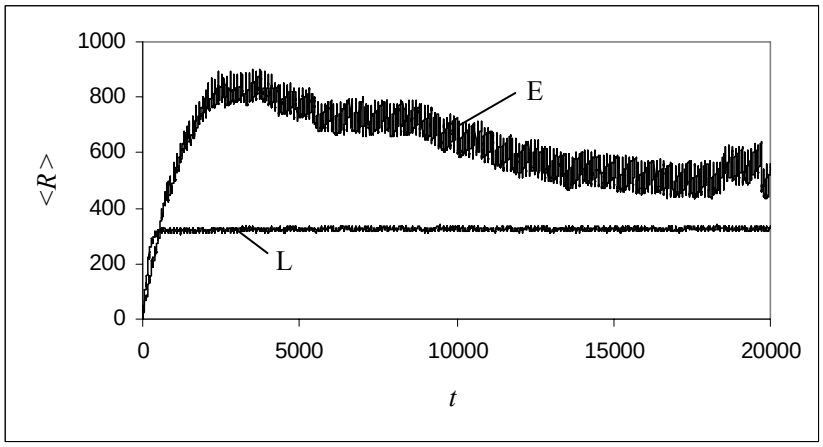

Рис. 4. Зависимости <*R*>(*t*) для случаев чистого обучения L и чистой эволюции E. Усреднено по 100 различным расчетам

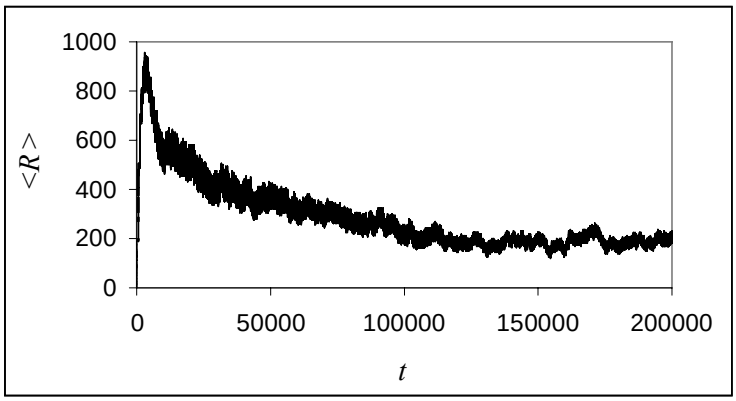

Рис. 5. Зависимость <*R*>(*t*) для случая чистой эволюции E. Усреднено по 100 различным расчетам

Объяснение полученным зависимостям состоит в том, что при чистой эволюции (случай E) важную роль играет размножение агентов (именно при размножении происходят отбор и мутации агентов, обеспечивающие оптимизацию их поведения), которое в расчетах становится достаточно частым не с самого начала эволюции популяции, а спустя некоторое время, при $t > 100000$. При этих временах агенты выполняют действие «деление» примерно в 40% тактов времени. При размножении агенты-родители отдают половину своего ресурса потомкам, поэтому средний ресурс агентов популяции <*R*> уменьшается. Если в компьютерной программе искусственно исключить деление ресурса $R(t)$ пополам между родителями и потомками при рождении новых агентов, то в случае E максимум в кривых <*R*>(*t*) исчезает. При обучении роль размножения невелика (агенты выполняют это действие всего в 3% тактов времени) и соответствующего уменьшения ресурса не происходит, поэтому в случаях L и LE установившийся средний ресурс агентов выше, чем в случае E.

При эволюционной оптимизации при $t \approx 200000$ (когда поведение агентов уже практически не менялось) действия агентов распределялись следующим образом. Действие «отдых» осуществляло небольшое число агентов (примерно 5% агентов популяции), действие «питание» – 55% агентов, действие «деление» – 40% агентов. При этом в моменты, непосредственно следующие за сменой статуса опасности клеток (5-10 тактов времени), доля агентов, выполнявших действие «деление», резко уменьшалась до 5% , доля агентов, выполнявших действия «отдых» и «питание», также уменьшалась, но всего на величину 2-5% от общего числа агентов популяции. В эти же моменты средняя по популяции частота действия «перемещение» (обеспечивающего перемещение агента в безопасную клетку) возрастала от почти нулевого значения до 40%.

При обучении (случай L) поведение агентов выходило на стационарный режим уже при $t \approx 5000$. В этом случае действия агентов распределялись следующим образом. Действие «отдых» осуществляли примерно 25% агентов популяции, действие «питание» – 70% агентов, действие «деление» – 3% агентов. При этом в моменты, непосредственно следующие за сменой статуса опасности клеток, частота действия «деление» практически не менялась, а частота выполнения агентами действий «отдых» и «питание» уменьшалась до 5% и 30%, соответственно. Частота действия «перемещение» сразу после смены статуса опасности клеток возрастала от 5% до 60%. Таким образом, динамика действий агентов в случаях E и L была сходна между собой. Отличие в основном состояло в том, что при эволюционной оптимизации существенно возрастала частота действия «деление» за счет других действий. В случае полной модели LE (обучение + эволюция) динамика частоты действий агентов только немного отличалась от таковой в случае чистого обучения L.

В изложенной модели не наблюдался эффект Болдуина (генетическая ассимиляция приобретаемых навыков), который был продемонстрирован в [7]. Это отличие связано с тем, что в данной модели при рождении потомка ресурс агента-родителя уменьшался, а в модели работы [7] при рождении новых агентов передачи ресурса от родителя к потомку не было. Можно ожидать, что эффект Болдуина должен наблюдаться или не наблюдаться в зависимости от величины ресурса, передаваемой от родителей к потомкам.

4. Выводы

Таким образом, построены и исследованы две модели автономных адаптивных агентов. В первой модели исследованы автономные агенты в двумерной клеточной среде, которые путем самообучения формируют свое поведение. Показано, что агенты в процессе обучения способны самостоятельно формировать последовательные цепочки действий, приводящие к росту ресурса агентов. Кроме того, если агент, наблюдая за ситуациями и выполняемыми действиями, самостоятельно оценивает усредненные частоты выполнения действий, то он способен самостоятельно формировать понятия, обобщающие сенсорную информацию.

Во второй модели продемонстрировано формирование достаточно естественного поведения агентов, обладающих потребностями питания, размножения, безопасности. Показано, что при эволюционной оптимизации систем управления агентов важную роль играет размножение. При этом, эволюционная оптимизация происходит медленней, чем при обучении. При объединении обучения с эволюционным поиском именно обучение играет основную роль, и результаты моделирования в случае объединенной модели близки к таковым в случае одного обучения.

Литература

1. Бесхлебнова Г.А., Редько В.Г. Модель формирования обобщенных понятий автономными агентами. – Четвертая международная конференция по когнитивной науке. – Томск, 22-26 июня 2010 г.: Тез. докл. в 2 томах: Томск: изд-во ТГУ, 2010. Т. 1. С. 174 – 175.
2. Редько В.Г. Моделирование когнитивной эволюции – перспективное направление исследований на стыке биологии и математики // *Математическая биология и биоинформатика (электронный журнал)*. – Т. 5. № 2, 2010. С. 215 – 229. URL: http://www.matbio.org/downloads/Redko2010(5_215).pdf
3. Редько В.Г., Бесхлебнова Г.А. Модель формирования адаптивного поведения автономных агентов – // Интегрированные модели и мягкие вычисления в искусственном интеллекте: V Международная научно-практическая конференция. – Коломна, 28-30 мая 2009 г. Сборник научных трудов в 2-х томах. М.: Физматлит, 2009. Т.1. С. 70 – 79.
4. Редько В.Г., Бесхлебнова Г.А. Моделирование адаптивного поведения автономных агентов // *Нейрокомпьютеры: разработка, применение*. – № 3, 2010. С. 33 – 38.
5. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. – М.: Эдиториал УРСС, 2002.
6. Holland J.H., Holyoak K.J., Nisbett R.E., Thagard P. Induction: Processes of Inference, Learning, and Discovery. – Cambridge: MIT Press, 1986.
7. Red'ko V.G., Mosalov O.P., Prokhorov D.V. A model of evolution and learning // *Neural Networks*. – V. 18. No. 5-6, 2005. PP. 738 – 745.
8. Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. – Cambridge: MIT Press, 1998.
9. 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*. – V. 11. No. 2, 2007. PP. 151 – 180.