Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: Материалы научной конференции. Том 1. Херсон: ХНТУ, 2010. С. 258-260.

МОДЕЛИ АВТОНОМНЫХ АДАПТИВНЫХ АГЕНТОВ, ОБЛАДАЮЩИХ ЭЛЕМЕНТАРНЫМИ КОГНИТИВНЫМИ СВОЙСТВАМИ

Г.А. Бесхлебнова, В.Г. Редько
Учреждение Российской академии наук Научно-исследовательский институт системных исследований РАН, Вавилова, 44/2, Москва, 119333, Россия, gab19@list.ru, vgredko@gmail.com

Ранее было предложено провести исследование когнитивной эволюции, эволюции познавательных способностей биологических организмов [1-3]. Это исследование целесообразно вести путем построения моделей последовательных этапов эволюции познавательных свойств – от простейших форм поведения организмов к логическим правилам, используемым в математических доказательствах. В настоящей работе характеризуются первые шаги по намеченному моделированию, излагаются две компьютерные модели адаптивного поведения автономных агентов, обладающих элементарными когнитивными свойствами.

В первой модели исследовалось поведение автономных агентов в двумерной клеточной среде. Каждый агент выполнял следующие действия: деление, питание, перемещение на одну клетку вперед, поворот направо или налево, нанесение удара по соседнему агенту, отдых. В части клеток двумерного мира имелись порции пищи. Агент обладал ресурсом $R(t)$, t – дискретное время. Ресурс $R(t)$ увеличивался при питании и уменьшался при выполнении агентом действий. При нанесении удара ударяющий агент отнимал ресурс у ударяемого.

Выбор действий агента обеспечивался имеющейся у него системой управления. Система управления агента представляла собой набор правил вида:

$$\mathbf{S}_k \rightarrow A_k \,, \tag{1}$$

где $\mathbf{S}_k$ и A_k – ситуация и действие, соответствующие этому правилу, k – номер правила. Ситуация $\mathbf{S}$ определялась наличием/отсутствием пищи или другого агента в поле зрения данного агента. Поле зрения включало 4 клетки: той клетки, в которой находился агент, клетки впереди агента и клеток справа/слева от агента. Каждое правило имело свой вес W_k , веса правил модифицировались при обучении агента. Начальный набор весов этих правил $\{W_{0k}\}$, получаемый агентом от родителя (с небольшими мутациями), представлял собой геном агента. В момент рождения агента текущие веса полагались равными начальным весам: $\{W_k\} = \{W_{0k}\}$. Таким образом, каждый агент имел 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.

Изменение весов W_k проводилось методом обучения с подкреплением [4] 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:

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

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.

При делении агента ресурс родителя делился пополам между родителем и потомком. Правила выбора действий потомка отличались от правил родителя малыми мутациями. При выборе агентом действия определялась текущая ситуация $\mathbf{S}(t)$ и из соответствующих ситуации правил (для которых $\mathbf{S}(t) = \mathbf{S}_k$) с вероятностью $1\text{-}\varepsilon$ выбиралось правило с максимальным весом и соответствующее выбранному правилу действие выполнялось ($0 < \varepsilon < 1$). С вероятностью ε выполнялось случайное действие. При моделировании использовался «метод отжига»: на начальных тактах моделирования, когда правила агентов еще не сформированы, полагалось $\varepsilon \sim 1$, т.е. была большая вероятность случайного выбора действий, а затем величина ε постепенно уменьшалась до нуля, и выбор действия осуществлялся в соответствии с правилами (1) и их весами.

Моделирование проводилось в рамках полной модели и в рамках упрощенной версии. В последнем случае изучалось обучение одного агента, у которого действия деление и борьба с другими агентами отсутствовали. В полной модели в процессе эволюции и обучения агентов формировалось следующее поведение: агенты преимущественно питались и часто отнимали ресурс друг у друга (нанося удары по соседям), изредка они выполняли и другие действия.

Результаты моделирования для упрощенной версии можно охарактеризовать следующим образом.

Каждая ситуация **S**(t) определялась 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. В конце расчета были выделены правила, имеющие достаточно большой вес. Этот набор правил можно рассматривать как обобщающие эвристики, формируемые агентом в процессе самообучения. Всего сформировалось 5 следующих эвристик: 1) если порция пищи расположена в той же клетке, в которой находится агент, то нужно выполнить действие «питание»; 2) если пищи нет в той клетке, в которой находится агент, и есть пища в клетке впереди агента, то нужно выполнить действие «перемещение вперед»; 3,4) если пищи нет ни в той клетке, в которой находится агент, ни в клетке впереди его, а есть пища в клетке справа/слева от агента, то нужно выполнить действие «поворот направо/налево», соответственно; 5) если вообще нет пищи в поле зрения агента, то нужно выполнить поисковое действие «перемещение вперед». Отметим, что действие «отдых» игнорировалось во всех ситуациях. Тем самым происходил отбор правил, приводящих к формированию цепочек действий агента, которые обеспечивали нахождение пищи и увеличение ресурса агента.

Во второй модели исследовалось адаптивное поведение автономных агентов, имеющих несколько естественных потребностей: питание, размножение, безопасность. Система управления агента была основана на правилах того же вида, что и в первой модели. Обучение проводилось методом обучения с подкреплением, согласно уравнению (2). Мир, в котором находились агенты, состоял из двух клеток: одна клетка являлась опасной для агентов, вторая – безопасной. Периодически статус клеток менялся: опасная клетка становилась безопасной, и, наоборот, клетка, бывшая безопасной, становилась опасной. Агент, находящийся в опасной клетке, каждый такт времени терял большой ресурс. В мире имелась восполняемая пища агентов. Агенты выполняли следующие действия: деление, питание, перемещение в другую (альтернативную из двух) клетку, отдых.

Специальным выбором параметров задавались следующие случаи: случай L (чистое обучение), в этом случае веса правил настраивались путем обучения с подкреплением; случай E (чистая эволюция), в этом случае веса правил модифицировались в результате мутаций и отбора делящихся агентов; случай LE (обучение + эволюция), для которого веса правил модифицировались как путем обучения, так и в процессе эволюционной оптимизации. Моделирование продемонстрировало, что во всех трех случаях после формирования правил агенты своевременно перемещались из опасной клетки в безопасную. При чистом обучении (случай L) агенты в основном выполняли действия, соответствующие потребностям питания и безопасности, а при эволюционной оптимизации (случай E) дополнительно к этому увеличивалась частота действий, соответствующих потребности размножения. В случае LE поведение агентов было близко к таковому в случае L. Итак, моделирование продемонстрировало формирование достаточно естественного поведения агентов. Существенно, что при эволюционной оптимизации важную роль играет размножение.

В изложенных моделях формировались элементарные когнитивные свойства агентов: они запоминали правила, определяющие их поведение.

Перспективы дальнейшего моделирования. Интересное направление развития моделей – введение мотиваций, количественно характеризующих потребности. Несложно ввести конкуренцию между мотивациями, что эквивалентно конкуренции между потребностями. Динамику отдельной мотивации целесообразно строить аналогично таковой в работе [5]. «Выигравшая» конкуренцию потребность использует свой блок структурированной системы управления агента. Блоки системы управления могут формироваться автоматически, в процессе обучения и эволюционной оптимизации поведения агентов. Оптимизация каждого блока должна происходить независимо от других блоков.

Далее имеет смысл ввести прогнозирование будущих ситуаций, следующих за намеченными действиями (методы прогнозирования известны). При наличии прогнозирования и обобщения ситуаций возможны дальнейшие шаги к использованию логических выводов при планировании автономными агентами своего поведения. Пример обобщения ситуаций – выделение наиболее важных черт сенсорной информации при формировании эвристик в первой из изложенных моделей. Кроме того, возможно использование методов семантического вывода [6]. Таким образом, имеются подходы к развитию моделей интеллектуальных и когнитивных свойств автономных агентов, к моделированию процессов возникновения простейших логических выводов.

ЛИТЕРАТУРА:

1. Редько В.Г. Актуальность моделирования когнитивной эволюции // «Научная сессия НИЯУ МИФИ-2010. Материалы избранных научных трудов по теме «Актуальные вопросы нейробиологии, нейроинформатики и когнитивных исследований». М.: НИЯУ МИФИ, 2010. С. 69-90.
2. Редько В.Г. Перспективы моделирования когнитивной эволюции // Третья международная конференция по когнитивной науке. Тезисы докладов в 2-х томах. Т. 2. М.: Художественно-издательский центр, 2008. С. 576-577.
3. Редько В.Г. Эволюция, нейронные сети, интеллект. – М.: КомКнига, 2005.
4. Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. – MIT Press, 1998.
5. 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.

6. Витяев Е.Е. Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов. – Новосибирск: НГУ, 2006.