

НАПРАВЛЕНИЕ ИССЛЕДОВАНИЙ «АДАПТИВНОЕ ПОВЕДЕНИЕ». СОСТОЯНИЕ И ПЕРСПЕКТИВЫ

В.Г. Редько

Учреждение Российской академии наук Научно-исследовательский институт системных исследований
РАН, Вавилова, 44/2, Москва, 119333, Россия, vcredko@gmail.com

В настоящей работе характеризуется направление исследований «Адаптивное поведение». Основной подход этого направления – конструирование и исследование модельных «организмов» (в виде компьютерной программы или робота), способных приспосабливаться к внешней среде. Работы в этом направлении часто рассматриваются как бионический подход к искусственному интеллекту. Далее кратко характеризуются модели адаптивного поведения, в том числе оригинальные. Обсуждается возможность будущих исследований когнитивной эволюции – эволюции познавательных способностей биологических и модельных автономных адаптивных организмов.

1. Общая характеристика направления «Адаптивное поведение»

Направление исследований «Адаптивное поведение» активно развивается с начала 1990-х годов [1]. Конструируемые и исследуемые в этом направлении искусственные «организмы» часто называются "аниматами" (от англ. animal + robot = animat) или автономными агентами. Исследователи адаптивного поведения стараются строить такие модели, которые применимы к описанию поведения как реального животного, так и искусственного анимата [2].

Ближняя цель направления «Адаптивное поведение» – исследовать архитектуры и принципы функционирования, которые позволяют животным или роботам жить и действовать в переменной внешней среде. Дальняя цель этого направления – попытаться проанализировать эволюцию когнитивных способностей животных и эволюционное происхождение человеческого интеллекта [3].

Хотя «официально» данное направление было провозглашено в 1990 году, были явные провозвестники этого направления. Например, хороший обзор ранних работ по адаптивному поведению представлен в книге [4].

В исследованиях адаптивного поведения используются методы вычислительного интеллекта:

- нейронные сети,
- генетический алгоритм [5] и другие методы эволюционной оптимизации,
- классифицирующие системы (Classifier Systems) [6],
- обучение с подкреплением (Reinforcement Learning) [7].

Отметим, что метод обучения с подкреплением – это метод самообучения автономного анимата или агента, который получает от учителя или из внешней среды только сигналы подкрепления, положительные (поощрения) или отрицательные (наказания), при этом учитель не объясняет обучаемому объекту, как именно нужно действовать. Этот метод имеет серьезную научную основу: он близок к методу динамического программирования, в нем используется теория марковских процессов.

2. Примеры моделей адаптивного поведения

2.1. Модели мозга и поведения в Институте нейронаук, руководимом Дж. Эдельманом. В Институте нейронаук (сайт института: <http://www.nsi.edu>) уже более 25 лет ведутся разработки поколений моделей работы мозга (Darwin I, Darwin II,...) и исследования адаптивного поведения искусственного организма NOMAD (Neurally Organized Mobile Adaptive Device); система управления этого устройства построена на базе таких моделей. Принципы моделирования NOMAD'a состоят в следующем:

- 1) устройство помещается в реальную физическую среду,
- 2) имеется некоторая поведенческая задача, которую должно решать устройство,
- 3) поведение устройства контролируется модельной нервной системой, которая отражает архитектуру мозга и динамику процессов в мозге,
- 4) поведение устройства и процессы в модельной нервной системе должны допускать сравнение с экспериментальными биологическими данными.

В одной из последних работ по NOMAD'у [8] промоделировано поведение мыши в лабиринте Морриса. Исследования поведения мыши или крысы в лабиринте Морриса состоят в следующем. Имеется бассейн с непрозрачной жидкостью, на бортах бассейна имеются рисунки, которые мышь видит и может использовать для ориентировки. В определенном месте бассейна расположена скрытая платформа, которую мышь может найти и тем самым спастись – не утонуть. Мышь бросают в бассейн, она плавает некоторое время и либо находит платформу и спасается, либо начинает тонуть (тогда ее спасает экспериментатор). После ряда экспериментов мышь начинает использовать ориентиры на бортах бассейна и находить платформу за достаточно короткое время.

Поведение NOMAD'a в лабиринте Морриса моделировалась следующим образом [8]. NOMAD представлял собой подвижное устройство на колесах, управляемое нейронной сетью, состоящей из 90000 нейронов, в которой было выделено 50 различных нейронных областей, в частности, были выделены несколько областей гиппокампа. Число синаптических контактов между нейронами равно $1.4 \cdot 10^6$. Программно нейронная сеть была реализована на основе компьютерного кластера.

Сенсорная система NOMAD'a включала зрение, обонятельную систему и специальный детектор скрытой от зрения платформы, позволяющий обнаруживать эту платформу только тогда, когда NOMAD находился непосредственно над ней.

NOMAD помещался в комнату, в которой была скрытая платформа; на стенах комнаты были разноцветные полосы – ориентиры. В начале каждого из компьютерных экспериментов NOMAD помещался в разные участки комнаты, задача NOMAD'a была найти скрытую платформу. Обучение нейронных сетей NOMAD'a осуществлялось на основе поощрений (получаемых при нахождении скрытой платформы) и наказаний (получаемых при приближении к стенам комнаты).

Было продемонстрировано, что

- 1) NOMAD обучается находить платформу достаточно быстро (за 10-20 попыток);
- 2) в модельном гиппокампе формируются нейроны места, активные только тогда, когда NOMAD находится в определенных участках комнаты;
- 3) в модельном гиппокампе формируются связи между отдельными нейронными областями, отражающие причинно-следственные зависимости.

Итак, изложенная модель представляет собой интересное эмпирическое компьютерное исследование самообучающегося адаптивного устройства, хорошо продуманное с биологической точки зрения.

2.2. Модель эволюционного возникновения коммуникаций в коллективе роботов. В работе [9] исследовались вопросы: Как могут эволюционно возникнуть коммуникации между модельными организмами? Как в эволюционном процессе может сформироваться сигнальная обработка информации?

Рассматривалась следующая проблема. Есть четыре робота, каждый из которых управляется рекуррентной нейронной сетью, состоящей из 5-ти нейронов. На входы нейронов поступают сигналы от 8-ми инфракрасных датчиков и от 4-х датчиков, воспринимающих звуковые сигналы с разных сторон. Нейронная сеть имела 3 выходных нейрона, два из которых определяли скорость движения двух колес робота, а третий нейрон – интенсивность силы звука, издаваемого в данный момент роботом. В ограниченной области пространства находились две кормушки, и роботам нужно было, используя свои нейронные сети и звуковые сигналы разной интенсивности, как можно быстрее распределиться по кормушкам: по два робота на каждую из кормушек.

Веса синапсов нейронных сетей роботов оптимизировались эволюционным путем. В результате в течение 2000 поколений у роботов сформировались сигналы 5-ти различных видов (т.е. разной интенсивности). Используя эти сигналы, роботы достаточно устойчиво находили требуемое распределение по кормушкам.

В целом в [9] продемонстрировано, что в эволюционирующей популяции роботов, управляемых рекуррентными нейронными сетями, может формироваться сигнальная система, позволяющая решать нетривиальную задачу распределения роботов по кормушкам.

2.3. Бионическая модель поискового адаптивного поведения. Одно из актуальных направлений исследований в рамках моделирования адаптивного поведения – имитация поискового поведения животных. В работе [10] исследовано поисковое поведение на примере личинок ручейников, обитающих на дне водоемов. Личинки носят на себе «домики» – трубку из песка и других частиц, которые они собирают на дне водоемов. Частицы скрепляются между собой по краям с помощью клейкой белковой нити. Строительство домика требует меньше времени, энергии и белка, если личинки используют относительно крупные и плоские частицы. Однако поиск крупных частиц на дне водоема требует затрат времени и энергии, не известных личинке заранее.

В [10] построена компьютерная модель поискового поведения личинок ручейников, строящих чехол-домик из частиц разного размера и ведущих поиск скоплений подходящих частиц. Модель использует понятие мотивации к прикреплению частиц к домику. Динамика регулирующей поведение мотивации $M(t)$ описывается уравнением:

$$M(t) = k_1 M(t-1) + \xi(t) + I(t), \quad (1)$$

время t дискретно, k_1 – параметр, характеризующий медленную релаксацию мотивации ($0 < k_1 < 1$, $1 - k_1 \ll 1$), $\xi(t)$ – случайные вариации мотивации, величина $I(t)$ характеризует направленное изменение мотивации при тестировании модельной личинкой частиц:

$$I(t) = k_2 (S_{curr} - S_{last}) / S_{last}, \quad (2)$$

где k_2 – положительный параметр, S_{curr} – площадь тестируемой в данный момент частицы, S_{last} – площадь последней протестированной ранее частицы. Если мотивация $M(t)$ достаточно велика, то происходит сбор и

прикрепление частиц к домику, если $M(t)$ мала, то модельная личинка ищет новое место с подходящими размерами частиц.

Построенная модель качественно согласуется с биологическими экспериментальными данными: как в эксперименте, так и в модели к домику преимущественно прикрепляются крупные частицы, есть сильный разброс числа прикрепляемых частиц и момента начала прикрепления, число прикрепляемых частиц в обоих случаях невелико.

Согласно (1), (2) динамика мотивации $M(t)$ проста и эффективна. Такая динамика может быть использована при моделировании регулирования переключений между тактиками поведения в случае нескольких потребностей и целей живого организма или искусственного анимата, например, мобильного робота.

Аналогичная динамика $M(t)$ может быть использована и в других подобных задачах. Например, ее можно использовать при поиске экстремума функции нескольких переменных. Такое моделирование было нами проведено и показана возможность реализации аналога известного «овражного» метода [11] оптимизации. При этом вводилась мотивация к сохранению направления поискового движения, при большой величине $M(t)$ направление движения сохраняется, при малых $M(t)$ направление поиска случайно варьируется. Если минимизируемая функция имеет глубокий «овраг», в котором она слабо меняется, то сначала в данном методе быстро находится сам овраг, а затем происходит постепенная минимизация функции при движении вдоль оврага.

2.4. Модель эволюции популяции самообучающихся агентов. В [12] исследовалась модель эволюции популяции самообучающихся автономных агентов. Модель отрабатывалась на примере агента-брокера. Задача агента была достаточно проста: надо было предсказывать курс акций, и при прогнозе роста курса вкладывать весь имеющийся у агента капитал в акции, при падении курса вкладывать капитал в деньги.

При моделировании считалось, что каждый такт времени агент выбирал одно из двух действий: 1) вложить весь свой капитал в акции, либо 2) вложить весь капитал в деньги. Система управления агента представляла собой простой нейросетевой адаптивный критик [13], состоящий из двух нейронных сетей (НС): Модель и Критик. НС Модель предназначена для прогнозирования изменения курса временного ряда. НС Критик предназначена для оценки ожидаемого прироста капитала в течение достаточно длительного времени в будущем для каждого из действий. Самообучение агента осуществлялось методом обучения с подкреплением [7], величина подкрепления в определенный такт времени была равна приросту капитала агента в данный такт времени. В процессе обучения веса синапсов НС подстраивались градиентным методом с целью минимизации ошибки предсказания и уточнения оценок ожидаемого прироста капитала.

Кроме обучения осуществлялась эволюционная оптимизация весов синапсов НС. Рассматривалась дарвиновская эволюция популяции n агентов. Продолжительность каждого поколения равна T тактов времени. Начальные веса синапсов обеих НС (Модели и Критика) формируют геном агента \mathbf{G} . Геном \mathbf{G} задается в момент рождения агента (в начале поколения) и не меняется в течение его жизни. Текущие веса синапсов НС \mathbf{W} подстраиваются в течение жизни агента путем обучения. В момент рождения агента $\mathbf{W} = \mathbf{G}$.

В конце каждого поколения определяется агент, прирост капитала которого в течение поколения максимален. Этот лучший агент порождает n потомков, которые составляют новое поколение. Геномы потомков \mathbf{G} отличаются от генома родителя небольшими мутациями.

Модель была реализована в виде компьютерной программы. Изучалась роль эволюции и обучения при формировании весов синапсов нейронных сетей агентов. Были проанализированы следующие варианты рассматриваемой модели:

- Случай L (чистое обучение), т.е. рассматривался отдельный агент, который только обучался;
- Случай E (чистая эволюция), т.е. рассматривалась эволюционирующая популяция без обучения;
- Случай LE (обучение + эволюция), т.е. полная модель.

Было проведено сравнение капитала (ресурса), приобретаемого агентами за 200 временных тактов для этих трех способов адаптации. Было продемонстрировано, что существует симбиотическое взаимодействие между обучением и эволюцией.

Если длительность поколения T была достаточно большой (1000 и более тактов времени), то для случая LE часто наблюдалось сильное влияние обучения на эволюционный процесс. В первых поколениях эволюционного процесса существенный рост ресурса агентов наблюдался не с самого начала поколения, а спустя 200-300 тактов, т.е. агенты явно обучались в течение своей жизни находить приемлемую стратегию поведения, и только после смены 3-5 поколений рост ресурса начинался с самого начала поколения. Это можно интерпретировать как проявление известного эффекта Болдуина [14, 15]: приобретаемый путем обучения навык в течение ряда поколений становился наследуемым. Итак, хотя эволюционный процесс и дарвиновский, тем не менее, в рамках эволюции наблюдаются особенности, характерные для ламарковской эволюции: приобретаемые навыки переходят в наследуемые, причем это происходит быстро, в течение всего 3-5 поколений эволюции.

3. Контуры программы будущих исследований когнитивной эволюции

Анализ моделей адаптивного поведения показывает, что хотя проделана большая работа и есть много интересных моделей, ученые еще далеки от понимания того, как возникали и развивались системы

управления живых организмов, как развитие этих систем способствовало эволюции когнитивных способностей животных, и как процесс такой когнитивной эволюции привел к возникновению логического мышления. В порядке обсуждения предложим контуры программы будущих исследований, нацеленных на моделирование когнитивной эволюции. Целесообразно в эти исследования включить следующие этапы.

А) Моделирование адаптивного поведения аниматов с несколькими естественными потребностями: питания, размножения, безопасности. Моделирование в этом направлении уже начато [16].

Б) Исследование перехода от физического уровня обработки информации в нервной системе животных к уровню обобщенных образов. Такой переход можно рассматривать, как появление в «сознании» животного свойства «понятие». Обобщенные образы можно представить как мысленные аналоги наших слов, не произносимые животными, но реально используемые ими. Использование понятий приводит к существенному сокращению требуемой памяти и времени обработки информации, поэтому оно должно быть эволюционно выгодным.

В) Исследование процессов формирования причинных связей в памяти животных. По-видимому, запоминание причинно-следственных связей между событиями во внешней среде и адекватное использование этих связей в поведении – одно из ключевых свойств активного познания животным закономерностей внешнего мира. Естественный следующий шаг – переход от отдельных причинных связей к логическим выводам на основе уже сформировавшихся знаний.

Г) Исследование процессов формирования логических выводов в «сознании» животных. Фактически, уже на базе классического условного рефлекса животные способны делать «логический вывод» вида: {УС, УС --> БС} => БС или «Если имеет место условный стимул (УС), и за условным стимулом следует безусловный, то нужно ожидать появления безусловного стимула (БС)». В определенной степени такие выводы подобны выводам математика, доказывающего теоремы. И целесообразно разобраться в системах подобных выводов, понять, насколько адаптивна логика поведения животных и насколько она подобна нашей, человеческой логике.

Перечисленные пункты очерчивают круг исследований от моделирования простейших форм поведения к логическим правилам, используемым в математике. Работы в этих направлениях уже ведутся, но четкой последовательности серьезных, канонических моделей еще нет.

ЛИТЕРАТУРА:

1. Meyer J.-A., Wilson S.W. (Eds.). From Animals to Animats. Proceedings of the First International Conference on Simulation of Adaptive Behavior. – Cambridge: MIT Press, 1991.
2. Непомнящих В.А. Поиск общих принципов адаптивного поведения живых организмов и аниматов // Новости искусственного интеллекта. 2002. № 2. С. 48-53.
3. Donnat J.Y., Meyer J.-A. Learning reactive and planning rules in a motivationally autonomous animat // IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 1996. V. 26. N. 3. PP. 381-395.
4. Гаазе-Рапопорт М.Г., Поспелов Д.А. От амебы до робота: модели поведения. – М.: Наука, 1987. М.: УРСС, 2004.
5. Holland J.H. Adaptation in Natural and Artificial Systems. – Ann Arbor, MI: The University of Michigan Press, 1975 (1st edn). Boston, MA: MIT Press., 1992 (2nd edn).
6. Holland J.H., Holyoak K.J., Nisbett R.E., Thagard P. Induction: Processes of Inference, Learning, and Discovery. – Cambridge, MA: MIT Press, 1986.
7. Sutton R., Barto A. Reinforcement Learning: An Introduction. – Cambridge: MIT Press, 1998.
8. Krichmar J.L., Seth A.K., Nitz D.A., Fleischer J.G., Edelman G.M. Spatial navigation and causal analysis in a brain-based device modeling cortical-hippocampal interactions // Neuroinformatics, 2005. V. 3. N. 3. PP. 197-221.
9. Marocco D., Nolfi S. Origins of communication in evolving robots // In: Nolfi S., Baldassarre G., Calabretta R., Hallam J., Marocco D., Miglino O., Meyer J.-A., Parisi D. (Eds.). From Animals to Animats 9: Proceedings of the Ninth International Conference on Simulation of Adaptive Behavior. – LNAI. Volume 4095. Berlin, Germany: Springer Verlag, 2006. PP. 789-803.
10. Непомнящих В.А., Попов Е.Е., Редько В.Г. Бионическая модель адаптивного поискового поведения // Изв. РАН. Теория и системы управления, 2008. № 1. С. 85-93.
11. Гельфанд И.М., Цетлин М.Л. Принцип нелокального поиска в задачах автоматической оптимизации // ДАН СССР, 1961. Т. 137. № 2. С. 295-298.
12. Red'ko V.G., Mosalov O.P., Prokhorov D.V. A model of evolution and learning // Neural Networks, 2005. V. 18. N. 5-6. PP. 738-745.
13. Редько В.Г., Прохоров Д.В. Нейросетевые адаптивные критики // Научная сессия МИФИ-2004. VI Всероссийская научно-техническая конференция «Нейроинформатика-2004». Сборник научных трудов. Часть 2. М.: МИФИ, 2004. С. 77-84.
14. Baldwin J.M. A new factor in evolution // American Naturalist, 1896. V. 30. N. 354. PP. 441-451.
15. Turney P., Whitley D., Anderson R. (Eds.). Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect // Special Issue of Evolutionary Computation on the Baldwin Effect. V. 4. N. 3, 1996.

16. Редько В.Г., Бесхлебнова Г.А. Модель формирования адаптивного поведения автономных агентов // «Интегрированные модели и мягкие вычисления в искусственном интеллекте». Сборник трудов V-й Международной научно-практической конференции. Т. 1. М.: Физматлит, 2009. С. 70-79.