

Каков естественный путь к искусственному интеллекту? *

В.Г.Редько

Институт оптико-нейронных технологий РАН

Аннотация. Обсуждаются подходы к теоретическим исследованиям когнитивной эволюции и аргументируется, что эти исследования могут составить естественнонаучную основу будущих разработок искусственного интеллекта.

В настоящей статье обсуждаются подходы к теоретическим исследованиям когнитивной эволюции — эволюции познавательных способностей биологических организмов — и аргументируется, что эти исследования могут составить естественнонаучную основу будущих разработок искусственного интеллекта (ИИ), характеризуются философские предпосылки исследований когнитивной эволюции, анализируется направление исследований «Адаптивное поведение» (анимат-подход к ИИ), которое может рассматриваться как задел моделей когнитивной эволюции, представлен проект «Мозг анимата», нацеленный на формирование общей платформы для моделирования адаптивного поведения, и предложены контуры программы будущих исследований когнитивной эволюции.

1. Философские предпосылки исследований когнитивной эволюции. Существует глубокая гносеологическая проблема: *почему человеческое мышление применимо к познанию природы?* Ведь далеко не очевидно, что те мыслительные процессы, которые мы используем в научном познании, применимы к процессам, происходящим в природе, так как эти два типа процессов различны. Рассмотрим, например, физику, наиболее фундаментальную из естественнонаучных дисциплин. Мощь физики связана с эффективным применением математики. Но математик строит свои теории совсем независимо от внешнего мира, используя свое мышление (в тиши кабинета, лежа на диване, в изолированной камере...). Почему же результаты, получаемые математиком, применимы к реальной природе?

Можно ли конструктивно подойти к решению этих вопросов? Скорее всего, да. Чтобы продемонстрировать такую возможность, будем рассуждать следующим образом.

Рассмотрим одно из элементарных правил, которое использует математик в логических заключениях, правило *modus ponens*: "если имеет место *A*, и из *A* следует *B*, то имеет место *B*", или $\{A, A \rightarrow B\} \Rightarrow B$.

А теперь перейдем от математика к собаке И.П. Павлова. Пусть у собаки вырабатывают условный рефлекс, в результате в памяти собаки формируется связь "за УС должен последовать БС" (УС — условный стимул, БС — безусловный стимул). И когда после выработки рефлекса собаке предъявляют УС, то она, помня о хранящейся в ее памяти "записи": УС \rightarrow БС, делает элементарный "вывод" $\{УС, УС \rightarrow БС\} \Rightarrow БС$. И у собаки, ожидающей БС (скажем, кусок мяса), начинают течь слюнки.

Конечно, применение правила *modus ponens* (чисто дедуктивное) математиком и индуктивный "вывод", который делает собака, явно различаются. Но можем мы ли думать об эволюционных корнях логических правил, используемых в математике? Да, вполне можем: умозаключение математика и индуктивный вывод собаки аналогичны.

Более того, было бы очень интересно попытаться построить модели эволюционного происхождения мышления. По-видимому, наиболее четкий путь такого исследования — построение математических и компьютерных моделей "интеллектуальных изобретений" биологической эволюции, таких, как безусловный рефлекс, привыкание (угасание реакции на биологически нейтральный стимул), условный рефлекс, цепи рефлексов, ..., логика. Иными словами, целесообразно с помощью моделей представить общую картину когнитивной

* Работа выполнена при финансовой поддержке РФФИ (проект № 04-01-00179).

эволюции: эволюции когнитивных способностей животных и эволюционного происхождения интеллекта человека.

2. Почему эти исследования актуальны

А) Эти исследования интересны с естественнонаучной точки зрения. Интеллект, мышление, логика — основа когнитивных, познавательных способностей человека. «Смотри в корень» — надо разобраться, как эти способности произошли.

Б) Исследования происхождения естественного интеллекта могут стать научной основой для разработок искусственного интеллекта.

В) Эти исследования важны с философской, эпистемологической точки зрения; они могут способствовать прояснению причин происхождения интеллекта человека и причин применимости человеческого мышления в познании природы.

Г) Эти исследования важны с социальной точки зрения; они должны способствовать развитию научного миропонимания, научной культуры человеческого сообщества, с их помощью возможно повышение престижа науки. Узнать ответ на вопрос: как произошел интеллект? — интересно и ученому, и просто любознательному человеку.

3. Модели адаптивного поведения — задел исследований когнитивной эволюции. С начала 1990-х годов за рубежом активно развивается направление исследований "Адаптивное поведение" (Редько, 2005; **от моделей поведения к искусственному интеллекту**, 2006; Meyer, Wilson, 1990). Основной подход этого направления — конструирование и исследование искусственных (в виде компьютерной программы или робота) "организмов", способных приспосабливаться к внешней среде. Эти организмы называются *аниматами* (от англ. animal + robot = animat). Данное направление исследований рассматривается как бионический подход к разработке систем искусственного интеллекта.

Программа-минимум направления "Адаптивное поведение" — исследовать архитектуры и принципы функционирования, которые позволяют животным или роботам жить и действовать в переменной внешней среде.

Программа-максимум этого направления — попытаться проанализировать эволюцию когнитивных способностей животных и эволюционное происхождение человеческого интеллекта.

Программа-максимум как раз соответствует задачам исследования когнитивной эволюции.

Хотя "официально" направление "Адаптивное поведение" было провозглашено в 1990 г., были явные провозвестники этого направления, например, в нашей стране еще в 1960–1970-х гг. исследования адаптивного поведения вели ряд талантливых ученых: М.Л. Цетлин, М.М. Бонгард, Д.А. Поспелов.

4. Исследователи адаптивного поведения. Исследования по адаптивному поведению ведутся в ряде университетов и лабораторий, таких как:

- AnimatLab (Париж, руководитель — один из инициаторов данного направления Жан-Аркадий Мейер). Подход AnimatLab (сайт лаборатории: <http://animatlab.lip6.fr/index.en.html>) предполагает, что система управления анимата может формироваться и модифицироваться посредством 1) *обучения*, 2) индивидуального *развития* (онтогенеза) и 3) *эволюции*.
- Лаборатория искусственного интеллекта в университете Цюриха (руководитель Рольф Пфейфер). Основной подход этой лаборатории (<http://www.ifi.unizh.ch/groups/ailab/>) — познание природы интеллекта путем его конструирования ("understanding by building"). Он включает в себя 1) построение моделей биологических систем, 2) исследование общих принципов естественного интеллекта животных и человека, 3) использование этих принципов при конструировании роботов и других искусственных интеллектуальных систем.

- Лаборатория искусственной жизни и роботики в Институте когнитивных наук и технологий (Рим, руководитель Стефано Нолфи), ведущая исследования в области эволюционной роботики и принципов формирования адаптивного поведения (<http://gral.ip.rm.cnr.it/>).
- Лаборатория информатики и искусственного интеллекта в Массачусетском технологическом институте (руководитель Родни Брукс), которая ведет исследования широкого спектра интеллектуальных и адаптивных систем, включая создание интеллектуальных роботов (<http://www.csail.mit.edu/>).
- Институт нейронаук Дж. Эдельмана, где ведутся разработки поколений моделей работы мозга (Darwin I, Darwin II,...) и исследования поведения искусственного организма NOMAD (Neurally Organized Mobile Adaptive Device), построенного на базе этих моделей (<http://www.nsi.edu/>).

В России исследования адаптивного поведения пока ведутся скромными усилиями ученых-энтузиастов, среди этих работ следует отметить (От моделей поведения к искусственному интеллекту, 2006):

- модели поискового адаптивного поведения (В.А. Непомнящих, Институт биологии внутренних вод им. И.Д. Папанина РАН);
- концепции и модели автономного адаптивного управления на основе аппарата эмоций (А.А. Жданов, Институт системного программирования РАН);
- разработку принципов построения систем управления антропоморфных и гуманоидных роботов (Л.А. Станкевич, Санкт-Петербургский политехнический университет);
- разработку нейросетевых моделей поведения роботов и робототехнических устройств (А.И. Самарин, НИИ нейрокибернетики им. А.Б. Когана РГУ);
- модели адаптивного поведения на основе эволюционных и нейросетевых методов (В.Г. Редько, М.С. Бурцев, О.П. Мосалов, Институт оптико-нейронных технологий РАН, Институт прикладной математики им. М.В. Келдыша РАН).

5. Проект "Мозг анимата". Анализ исследований в рамках анимат-подхода показывает, что разработанные модели пока еще очень фрагментарны и иллюстрируют только отдельные стороны адаптивного поведения. Поэтому в наших работах (Редько, 2005; От моделей поведения к искусственному интеллекту, 2006; Red'ko, Prokhorov, Burtsev, 2004) предложен проект "Мозг анимата", который нацелен на формирование общей схемы построения таких моделей. Проект основан на теории функциональных систем П.К. Анохина и предполагает, что система управления анимата обеспечивает целенаправленное адаптивное поведение, соответствующее естественным потребностям: питания, размножения, безопасности, накопления знаний. Система управления имеет блочную архитектуру, каждый блок представляет собой функциональную систему (ФС). Управление может передаваться от одних ФС к другим. В каждый момент активна только одна ФС, в которой формируется текущее действие.

Каждая ФС содержит две нейронные сети: Контроллер и Модель. На вход активной ФС поступает информация о текущей ситуации $S(t)$. $S(t)$ — набор параметров (сенсорных сигналов), характеризующий внешнюю и внутреннюю среду анимата (время t дискретно, $t = 1, 2, \dots$). Контроллер по известной ситуации $S(t)$ формирует действие анимата $A(t)$. Модель по известным векторам $S(t)$ и $A(t)$ прогнозирует ситуацию в следующий такт времени $S(t+1)$. После выполнения действия происходит передача управления той ФС, которая будет активна в следующий момент времени $t+1$. Передача управления происходит в соответствии со связями между ФС.

Анимат обучается без учителя, в результате непосредственного взаимодействия с внешней средой. Обучение в нейронных сетях осуществляется следующим образом. Имеется два режима обучения: 1) грубого поиска, 2) тонкой доводки.

Грубый поиск происходит, если есть сильное рассогласование между ожидаемым и полученным результатом: прогноз ситуации $S_{pr}(t+1)$ существенно отличается от реальной ситуации $S(t+1)$. При грубом поиске происходит реорганизация работы системы управления анимата: модифицируются связи между ФС, а также происходит случайная генерация новых ФС, и среди этих ФС происходит отбор тех, которые в наибольшей степени способны выполнять действия, направленные на достижение текущей цели/подцели.

В режиме тонкой доводки обучение происходит путем настройки весов синапсов нейронов в ФС, активной в текущий момент времени, и в ФС, бывших активными несколько предыдущих тактов времени. При этом усиливаются/ослабляются связи в нейронных сетях Контроллеров, приведшие к положительным/отрицательным подкреплениям. Также происходит уточнение прогнозов, формируемых в нейронных сетях Моделей. Так как модификация весов синапсов происходит и в тех ФС, которые были активны в предыдущие такты времени, данная схема позволяет формировать цепочки последовательных действий.

Предполагается, что есть первичный и вторичный репертуар поведения. Первичный репертуар формируется эволюционным путем, путем селекции и мутаций ФС и связей между ними, а изложенное выше обучение формирует вторичный репертуар.

Проводится компьютерное моделирование с целью проверки эффективности прогнозов в рассматриваемой системе управления и исследования условий согласованности между формированием цепочек действий (обеспечиваемых Контроллерами) и познанием причинно-следственных закономерностей (обеспечиваемых Моделями).

6. Моделирование когнитивной эволюции — естественнонаучная основа разработок искусственного интеллекта (ИИ). Направление исследований ИИ, скорее всего, можно рассматривать как прикладное — применение принципов естественного интеллекта в искусственных практически важных для человека компьютерных системах. Судьба прикладных разработок зависит от наличия серьезного научного фундамента, на котором базируются такие разработки. Например, научной базой развития микроэлектроники во второй половине XX в. была физика твердого тела. При этом для физиков чисто научные исследования твердого тела были интересны фактически независимо от применения их исследований, в результате чего научная основа микроэлектроники интенсивно развивалась. И результаты микроэлектроники как наукоемкой технологии впечатляющи.

Моделирование когнитивной эволюции чрезвычайно интересно и важно с точки зрения научного миропонимания. Следовательно, можно ожидать, что такие исследования, которые способны представить картину эволюционного формирования естественного интеллекта, будут очень интересны для ученых. Но эти исследования могут быть тесно связаны и с разработками ИИ. И, следовательно, возможно взаимное обогащение фундаментальных и прикладных исследований природы интеллекта. Тем самым исследования когнитивной эволюции могли бы служить научной основой разработок систем ИИ.

7. Контуры программы будущих исследований. Анализ современных исследований адаптивного поведения показывает, что хотя проделана большая работа, ученые еще очень далеки от понимания того, как возникали и развивались системы управления живых организмов, как развитие этих систем сопровождалось эволюцией когнитивных способностей животных и как процесс когнитивной эволюции привел к возникновению интеллекта человека. Следовательно, есть огромная область чрезвычайно интересных исследований, которые только-только начинаются. Предложим эскизный план будущих исследований когнитивной эволюции.

а) Разработка схем и моделей адаптивного поведения на базе проекта "Мозг анимата". Воплощение в конкретные модели конструкций Мозга анимата разумно начать с анализа целостного адаптивного поведения простых «организмов», имеющих естественные потребности: питания, размножения, безопасности. Подход к такому моделированию

очерчен выше.

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 произносимых животными, но реально используемых ими. Например, у собаки явно есть понятия "хозяин", "свой", "чужой", "пища". И важно осмыслить, как такой весьма нетривиальный переход мог возникнуть в процессе эволюции.

в) Исследование процессов формирования причинной связи в памяти животных. Повидимому, 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. Такая связь формируется, например, при выработке условного рефлекса: животное запоминает связь между условным стимулом (УС) и следующим за ним безусловным стимулом (БС), что позволяет ему предвидеть события в окружающем мире и адекватно использовать это предвидение.

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: $\{УС, УС \rightarrow БС\} \Rightarrow БС$ или «Если имеет место условный стимул, и за условным стимулом следует безусловный, то нужно ожидать появления безусловного стимула». Можно даже говорить, что такие выводы подобны выводам математика, доказывающего теоремы (п. 1). И целесообразно разобраться в системах подобных выводов, понять, насколько адаптивна логика поведения животных и насколько она подобна нашей, человеческой логике.

д) Исследование коммуникаций, возникновения языка. Наше мышление тесно связано с языком, с языковым общением людей. Поэтому целесообразно проанализировать, как в процессе биологической эволюции возникал язык общения животных, как развитие коммуникаций привело к современному языку человека, как развитие коммуникаций и языка способствовало развитию логики, мышления, интеллекта человека.

Конечно же, перечисленные пункты формируют только контуры плана будущих исследований. Тем не менее уже сейчас видно, сколь широк фронт исследований и как много нетривиальной, интересной и важной работы предстоит сделать.

Литература

Редько В.Г. Эволюция, нейронные сети, интеллект: Модели и концепции эволюционной кибернетики. Серия "Синергетика: от прошлого к будущему". М.: УРСС, 2005. 224 с.

От моделей поведения к искусственному интеллекту. Серия "Науки об искусственном" / Под ред. В.Г. Редько. М.: УРСС, 2006. 456 с.

Meyer J.-A., Wilson S. W. Editors, From Animals to Animats. Proc. of the First International Conference on Simulation of Adaptive Behavior. The MIT Press: Cambridge, Massachusetts, London, England, 1990.

Red'ko V.G., Prokhorov D.V., Burtsev M.S. Theory of functional systems, adaptive critics and neural networks // Proc. International Joint Conference on Neural Networks (IJCNN 2004) Budapest, 2004. P.1787–1792.