FROM ANIMAL TO ANIMAT – НАПРАВЛЕНИЕ ИССЛЕДОВАНИЙ "АДАПТИВНОЕ ПОВЕДЕНИЕ"*

В.Г. Редько
Институт оптико-нейронных технологий РАН
E-mail: redko@iont.ru

Аннотация. В настоящей главе представлен аналитический обзор направления исследований "Адаптивное поведение", цель которого – исследовать архитектуры и принципы функционирования, позволяющие аниматам (модельным организмам) приспосабливаться к переменной внешней среде. Особое внимание уделяется работам Массачусетского университета (исследования Р. Саттона и Э. Барто по методу обучения с подкреплением) и работам AnimatLab (Париж) по анализу формирования архитектур управления на основе симбиоза 1) индивидуального обучения, 2) онтогенетического развития анимата, и 3) эволюционной оптимизации. Глава также включает изложение моделей эволюционного возникновения целенаправленного адаптивного поведения и проекта «Мозг Анимата», нацеленного на формирование общей схемы широкого спектра моделей адаптивного поведения, а также обсуждение подходов к исследованию проблемы происхождения интеллекта.

В процессе биологической эволюции возникли чрезвычайно сложные и вместе с тем удивительно эффективно функционирующие живые организмы. Эффективность, гармоничность и согласованность работы "компонент" живых существ обеспечивается биологическими управляющими системами.

Но каковы эти управляющие системы? Какие информационные процессы обеспечивают их работу? Какие архитектуры и принципы функционирования биологических систем управления обеспечивают способность животных приспосабливаться, адаптироваться к постоянно меняющимся условиям во внешней среде? До какой степени исследования биологических адаптивных систем могут способствовать развитию информационных технологий?

В настоящей главе речь пойдет об исследованиях, направленных на изучение этих интригующих проблем, а именно о современных исследованиях в области моделирования адаптивного поведения. Это направление исследований сложилось сравнительно недавно и основная его цель – изучение систем управления адаптивным поведением живого или искусственного организма (реализованного в виде робота или компьютерной модели). В 1-м разделе мы охарактеризуем в целом направление исследований "Адаптивное поведение", а также тесно связанное с ним направление "Искусственная жизнь". В разделе 2 характеризуются компьютерные методы, используемые в этих исследованиях, в частности, описывается метод обучения с подкреплением (reinforcement learning), разработанный Р. Саттоном и Э. Барто (Массачусетский университет), а также основанные на этом методе схемы адаптивного управления – нейросетевые адаптивные критики. В разделе 3 излагаются подходы к исследованиям адаптивного поведения, разрабатываемые в AnimatLab – одной из ведущих лабораторий в этой области. Далее (раздел 4) обсуждается общий концептуальный подход, базирующийся на теории функциональных систем П.К. Анохина, и который естественно использовать в будущих исследованиях. Затем следует описание конкретных моделей эволюционного возникновения целенаправленного адаптивного поведения (раздел 5). В разделе 6 излагается основанный на теории функциональных систем проект «Мозг Анимата», который может рассматриваться как общая платформа для систематического построения моделей адаптивного поведения. И в заключение (раздел 7) обсуждаются подходы к исследованию исключительно интересной и важной с научной точки зрения проблемы – проблемы происхождения интеллекта.

* Работа выполнена финансовой поддержке программы Президиума РАН "Интеллектуальные компьютерные системы" (проект 2-45) и РФФИ (проект № 04-01-00179).

1. Направления исследований "Адаптивное поведение" и "Искусственная жизнь"

В конце 1980-х - начале 1990-х годов возникли два интересных, тесно связанных между собой направления исследований: "Искусственная жизнь" (английское название Artificial Life или ALife) [1,2] и "Адаптивное поведение" (Adaptive Behavior) [3].

Первая конференция по **Искусственной жизни** состоялась в 1987 году в Лос Аламосе. Как сказал руководитель этой конференции К. Ленгтон, "основное предположение искусственной жизни состоит в том, что «логическая форма» организма может быть отделена от материальной основы его конструкции". Основной мотивацией исследований Искусственной жизни (ИЖ) служит желание понять и промоделировать формальные принципы организации биологической жизни.

Отметим, что хотя лозунг "Искусственная жизнь" был провозглашен в конце 1980-х, в действительности идейно близкие модели разрабатывались в 1950-70-е годы. Приведем два примера из истории отечественной науки.

В 1960-х годах блестящий кибернетик и математик М.Л. Цетлин предложил и исследовал модели автоматов, способных адаптивно приспосабливаться к окружающей среде. Работы М.Л. Цетлина инициировали целое научное направление, получившее название "коллективное поведение автоматов" [4,5].

В 1970-х годах под руководством талантливого кибернетика М.М. Бонгарда был предложен весьма нетривиальный проект "Животное", характеризующий адаптивное поведение искусственных организмов [6,7].

Типичные примеры современных моделей "Искусственной жизни" кратко охарактеризованы в [8].

Первую международную конференцию по Адаптивному поведению организовали Жан-Аркадий Мейер и Стюарт Вильсон в 1990 году в Париже.

Основной подход направления "**Адаптивное поведение**" – конструирование и исследование искусственных (в виде компьютерной программы или робота) "организмов", способных приспосабливаться к внешней среде [3, 9-11]. Эти организмы называются "*аниматами*" (от англ. animal + robot = animat). Часто используют также близкий термин "агент", подразумевая под этим термином модельный искусственный организм.

Поведение аниматов имитирует поведение животных. Исследователи направления "Адаптивное поведение" (АП) стараются строить такие модели, которые применимы к описанию поведения *как реального животного, так и искусственного анимата* [10,11].

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 [12].

Как и для ИЖ, для исследований АП характерен *синтетический подход*: здесь конструируются архитектуры, обеспечивающие "интеллектуальное" поведение аниматов. Причем это конструирование проводится как бы с точки зрения инженера: исследователь сам "изобретает" архитектуры, подразумевая, конечно, что какие-то подобные структуры, обеспечивающие адаптивное поведение, должны быть у реальных животных.

И так же, как для ИЖ, для направления АП были явные провозвестники этого направления до его "официального провозглашения". Яркий пример – хороший обзор ранних работ по адаптивному поведению, представленный в книге М.Г. Гаазе-Рапопорта, Д.А. Поспелова "От амебы до робота: модели поведения" [7].

Ряд современных моделей АП представлен в обзорах В.А. Непомнящих [10,11].

ИЖ и АП имеют много общего: синтетический подход к конструированию жизнеподобных "организмов", попытка промоделировать формальные законы жизни и систем управления, ориентация на компьютерные и математические модели, использование эволюционных концепций и моделей.

Эти направления используют ряд нетривиальных компьютерных методов:
- нейронные сети,
- генетический алгоритм [13] и другие методы эволюционной оптимизации,
- классифицирующие системы (Classifier Systems) [14],
- обучение с подкреплением (Reinforcement Learning) [15, 16].

Подчеркнем, что АП и ИЖ – активно развивающиеся направления исследований. По этим направлениям регулярно проводятся международные и европейские конференции "Simulation of Adaptive Behavior (From Animal to Animat)", "Artificial Life", "European Conference on Artificial Life". Издаются журналы "Adaptive Behavior" и "Artificial Life".

В целом соотношение между направлениями "Искусственная жизнь" и "Адаптивное поведение", используемыми в них компьютерными методами, их научным значением и их потенциальными применениями иллюстрируется рис. 1.

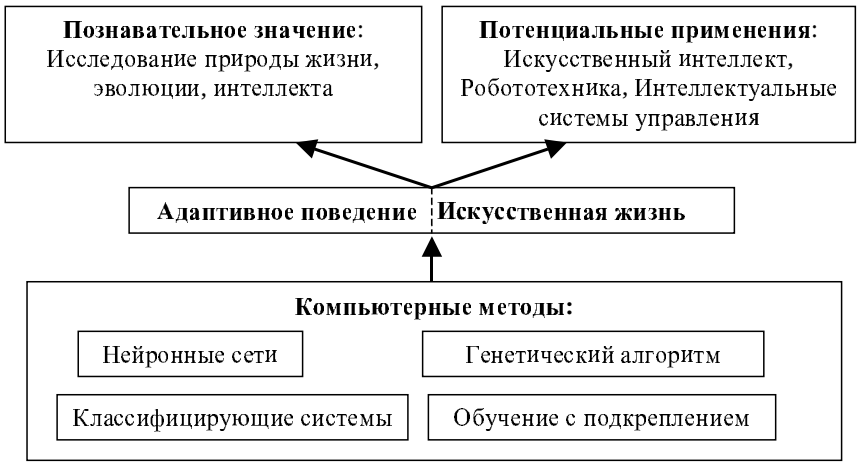


Рис. 1. Направления исследований "Адаптивное поведение" и "Искусственная жизнь".

Отметим, что направление исследований "Адаптивное поведение" выглядит значительно более серьезным, чем "Искусственная жизнь". В работах по ИЖ слишком много "игрушечности", исследователи часто просто играют с теми моделями, которые они придумывают, многих исследователей, возможно, привлекает красивый лозунг "Искусственная жизнь" (Хотя необходимо отметить, что на основе исследований искусственной жизни формируется новое направление прикладных разработок – многоагентное моделирование [17]). Исследования по адаптивному поведению более целенаправленны – они ориентированы на определенную и важную задачу – исследование систем управления, обеспечивающих поведение естественных или искусственных организмов. Более того, схемы управления адаптивным поведением тестируются на модельных (реализованных в компьютерных программах) или реальных роботах, что придает направлению исследований "Адаптивное поведение" определенность и надежность.

Подчеркнем, что в "Адаптивном поведении", как и в "Искусственной жизни", в основном используется *феноменологический подход* к исследованиям систем управления адаптивным поведением. Т.е. предполагается, что существуют формальные правила адаптивного поведения, и эти правила не обязательно связаны с конкретными микроскопическими нейронными или молекулярными структурами, которые есть у живых организмов. Скорее всего, такой феноменологический подход для исследований адаптивного поведения вполне имеет право на существование. В пользу этого тезиса приведем аналогию из физики. Есть термодинамика, и есть статистическая физика. Термодинамика описывает явления на феноменологическом уровне, статистическая физика характеризует те же явления на микроскопическом уровне. В физике термодинамическое и стат-физическое описания относительно независимы друг от друга, и вместе с тем, взаимодополнительны. По-видимому, и для описания живых организмов может быть аналогичное соотношение между феноменологическим (на уровне поведения) и микроскопическим (на уровне нейронов и молекул) подходами. При этом, естественно ожидать, что для исследования систем управления адаптивным поведением феноменологический подход должен быть более эффективен (по крайней мере, на начальных этапах работ), так как очень трудно сформировать целостную картину поведения на основе анализа всего сложного многообразия функционирования нейронов, синапсов, молекул.

2. Компьютерные методы

Обратим внимание на компьютерные методы, используемые в работах по ИЖ и АП (рис. 1).

Исследования по нейронным сетям и генетическому алгоритму достаточно хорошо известны отечественным ученым и поэтому мы охарактеризуем их только в общих чертах. Работы по классифицирующим системам и обучению с подкреплением практически не представлены в русскоязычной научной литературе, поэтому мы остановимся на этих работах немного подробней.

2.1. Нейронные сети

Современные исследования нейронных сетей чрезвычайно интересны, активно развиваются, здесь есть множество различных подходов, концепций и моделей. Большинство моделей основывается на схемах формальных нейронов, которые можно рассматривать как развитие схем нейронов, предложенных в пионерской работе У.С. Мак-Каллока и У. Питтса (1943 год) [18]. Формальный нейрон представляет собой пороговый элемент, на входах которого имеются возбуждающие и тормозящие синапсы, в нейроне определяется взвешенная сумма (с учетом весов синапсов) входных сигналов, при превышении этой суммой порога нейрона вырабатывается выходной сигнал.

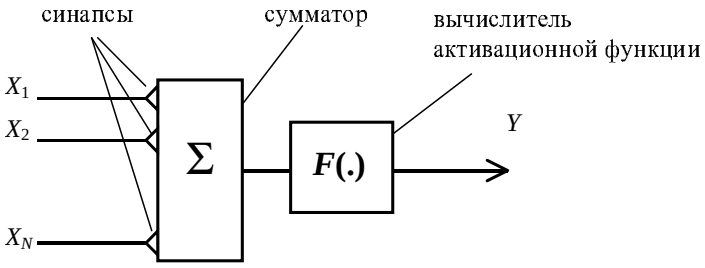


Рис. 2. Схема формального нейрона. X_i – входные сигналы, Y – выходной сигнал нейрона, $F(.)$ – активационная функция.

В общем виде работа формального нейрона (рис. 2) может быть описана уравнениями:

$Y_j = F(net_j - K_j)$,
$net_j = \Sigma_i w_{ji} X_i$,

где j – номер нейрона в сети, X_i – входные сигналы, Y_j – выходной сигнал нейрона, w_{ji} – веса синапсов, net_j – суммарное входное воздействие на нейрон, K_j – порог нейрона, $F(.)$ – активационная функция.

Активационная функция характеризует реакцию нейрона на входное воздействие net_j , она может быть пороговой:

$$F(a) = \begin{cases} 0 \text{ при a} < 0, \\ 1 \text{ при a} > 0, \end{cases}$$

или некоторой непрерывной, например, линейной:

$F(a) = ka$

или логистической:

$F(a) = 1/[1+exp(-a)]$.

В зависимости от реализуемого алгоритма на допустимые значения входов и выходов нейрона накладываются определенные ограничения: значения X_i и Y_j могут бинарными (т.е. равными 0 или 1), бинарными биполярными (+1 или -1), действительными и принадлежащими интервалу (0,1), действительными неотрицательными или произвольными действительными числами. Аналогичные ограничения накладываются на веса синапсов нейронов w_{ij} .

История исследования нейронных сетей испытывала взлеты и падения. Первый всплеск энтузиазма был в 1950-60-х годах. Его можно связать с работами Дж. фон Неймана по сравнительному анализу работы биологических нейронных сетей и компьютеров [19] и по разработке принципов построения надежных вычислительных систем из ненадежных компонент (фактически из формальных нейронов) [20], а также с основополагающими работами Ф. Розенблата по персептронам [21].

Второй нейросетевой бум возник в 1980-х годах. Во второй половине 80-х годов был предложен целый ряд интересных и содержательных моделей нейронных сетей. В этих моделях построены нейросети, выполняющие различные алгоритмы обработки информации: ассоциативная память [22-26], категоризация, т.е. разбиение множества образов на кластеры, состоящие из подобных друг другу [27], топологически корректное картирование [28], распознавание зрительных образов, инвариантное относительно деформаций и сдвигов в пространстве [29], решение задач комбинаторной оптимизации [30]. В большинстве моделей запоминание информации в нейронной сети (обучение) происходит в результате формирования весов синапсов нейронов. Во многих случаях это интерпретируется как формализация гипотезы Хебба [31], в соответствии с которой изменение состояния произвольного синапса определяется его текущим состоянием и активностью пре- и постсинаптических нейронов.

Один из важных и наиболее исследованных способов обучения нейронных сетей – метод обратного распространения ошибок [26], в котором формирование весов нейронной сети осуществляется путем оптимизации весов синапсов методом градиентного спуска.

В 1990-х годах активность по предложению новых архитектур и моделей нейронных сетей несколько снизилась, но зато нейросети (реализованные в компьютерных программах) и нейрочипы (специализированные микроэлектронные схемы, реализующие работу нейронных сетей) вошли в инженерный обиход.

2.2. Генетический алгоритм

Генетический алгоритм (ГА) [13, 32-34] – это компьютерная модель эволюции популяции искусственных "особей". Каждая особь характеризуется своей хромосомой $\boldsymbol{S_k}$, хромосома есть "геном" особи. Хромосома определяет приспособленность особи $f(\boldsymbol{S_k})$; $k = 1,..., n$; n – численность популяции. Хромосома есть цепочка символов $\boldsymbol{S_k} = (S_{k1}, S_{k2},...,S_{kN})$, N – длина цепочки. Символы выбираются из некоторого алфавита и интерпретируются как "гены" особи, расположенные в хромосоме $\boldsymbol{S_k}$. Задача алгоритма состоит в максимизации функции приспособленности $f(\boldsymbol{S_k})$.

Эволюция состоит из последовательности поколений. В каждом поколении отбираются особи с большими значениями приспособленностями. Хромосомы отобранных особей рекомбинируются и подвергаются малым мутациям. Формально, схема ГА может быть представлена следующим образом (популяция t-го поколения обозначается как $\{\boldsymbol{S_k}(t)\}$):

Шаг 0 . Создать случайную начальную популяцию $\{\boldsymbol{S_k}(0)\}$.
Шаг 1. Вычислить приспособленность $f(\boldsymbol{S_k})$ каждой особи $\boldsymbol{S_k}$ популяции $\{\boldsymbol{S_k}(t)\}$.
Шаг 2. Производя отбор особей $\boldsymbol{S_k}$ в соответствии с их приспособленностями $f(\boldsymbol{S_k})$ и применяя генетические операторы (рекомбинации и точечные мутации) к отобранным особям, сформировать популяцию следующего поколения $\{\boldsymbol{S_k}(t+1)\}$.
Шаг 3. Повторять шаги 1, 2 для $t = 1, 2, ...$, до тех пор, пока не выполнится некоторое условие окончания эволюционного поиска (прекращается рост максимальной приспособленности в популяции, число поколений t достигает заданного предела и т.п.).

Имеется ряд конкретных вариантов генетического алгоритма, которые отличаются по схемам отбора, рекомбинаций, по форме представления хромосом и т.д. Подробнее о ГА см., например, [8, 34]. В целом генетический алгоритм можно рассматривать, как простой эвристический метод оптимизации функций дискретных переменных.

Отметим, что в моделях адаптивного поведения и искусственной жизни часто не вводится функция приспособленности, а явно применяются только генетические операторы. Приспособленность проявляется естественным путем: особи рождаются, когда их родители готовы дать потомков, и погибают, когда не хватает пищи или когда их убивает и съедает хищник. В этом случае – при отсутствии явной функции приспособленности – говорят, что приспособленность эндогенна. Пример использования такой эндогенной приспособленности описан ниже в разделе 5.

Следует подчеркнуть, что генетический алгоритм – наиболее известный представитель целого семейства методов эволюционного моделирования, к которым можно также отнести:

- эволюционное программирование, ориентированное на оптимизацию непрерывных функций без использования рекомбинаций;
- эволюционную стратегию, ориентированную на оптимизацию непрерывных функций с использованием рекомбинаций;
- генетическое программирование, использующее эволюционный метод для оптимизации компьютерных программ [35, 36].

2.3. Классифицирующие системы

Классифицирующие системы (Classifier Systems) [14] предназначены для формирования правил поведения анимата.

Правила имеют вид:

Если <УСЛОВИЕ(Я)> , то <ДЕЙСТВИЕ> .

Условия представляют собой цепочки бинарных символов:

S_k = (1, 0, 1, 0, 0, 1).

Классифицирующая система содержит множество таких правил (классификаторов). Каждый классификатор имеет свой вес W_k (силу классификатора). При этом классификаторы допускают неопределенность – неполное совпадение всех символов в условиях. Т.е., условие может иметь вид:

S_k = (1, 0, #, 0, #, 1), # – любой символ из множества {0,1}.

Часть действий состоит в формировании условий для системы классификаторов, т.е. в процессе работы возможно образование цепочки последовательных действий.

Популяция классификаторов оптимизируется в результате двух процессов: обучения и эволюции.

При обучении модифицируются веса классификаторов W_k. Обучение происходит так называемым методом "пожарной бригады": при успехе поощряется не только тот классификатор, который привел к успешному действию, но и его предшественники.

В процессе эволюции с помощью генетического алгоритма формируются новые классификаторы.

Современное состояние исследований классифицирующих систем хорошо отражено в работе [37].

2.4. Обучение с подкреплением

Теория обучения с подкреплением (reinforcement learning) была развита в работах Р. Саттона и Э. Барто (Массачусетский университет).

Идейным вдохновителем этих работ был А.Г. Клопф (Air Force, USA), который в книге "Целеустремленный нейрон" предложил несколько спорную, но достаточно четкую и последовательную методологию исследований памяти, обучения, адаптивного поведения [38].

Общая схема обучения с подкреплением [15] показана на рис. 3.

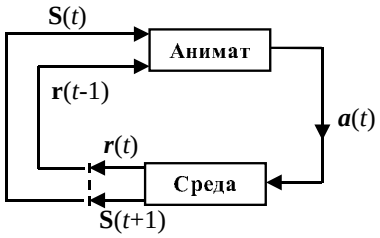


Рис.3. Схема обучения с подкреплением.

Рассматривается анимат, взаимодействующий с внешней средой. Время предполагается дискретным: t = 1,2,… В текущей ситуации анимат $\mathbf{S}(t)$ выполняет действие $a(t)$, получает подкрепление $r(t)$ и попадает в следующую ситуацию $\mathbf{S}(t+1)$. Подкрепление может быть положительным (награда) или отрицательным (наказание).

Цель анимата – максимизировать суммарную награду, которую можно получить в будущем в течение длительного периода времени. Подразумевается, что анимат может иметь свою внутреннюю "субъективную" оценку суммарной награды и в процессе обучения постоянно совершенствует эту оценку. Эта оценка определяется с учетом коэффициента забывания:

$$U(t) = \sum_{k=0}^{\infty} \gamma^k r(t+k), \qquad (1)$$

где $U(t)$ - оценка суммарной награды, ожидаемой после момента времени t, , γ – коэффициент забывания (дисконтный фактор), $0 < \gamma < 1$. Коэффициент забывания учитывает, что чем дальше анимат «заглядывает» в будущее, тем меньше у него уверенность в оценке награды («рубль сегодня стоит больше, чем рубль завтра»).

В процессе обучения анимат формирует *политику* (стратегию поведения). Политика определяет выбор (детерминированный или вероятностный) действия в зависимости от ситуации. Р. Саттон и Э. Барто [15] исследовали ряд методов формирования политики, основанных на динамическом программировании и теории марковских процессов.

Удалено: м

2.4.1. Метод SARSA

Если множество возможных ситуаций $\{\mathbf{S}_i\}$ и действий $\{a_j\}$ конечно, то существует простой метод обучения SARSA, каждый шаг которого соответствует цепочке событий $\mathbf{S}(t) \rightarrow a(t) \rightarrow r(t) \rightarrow \mathbf{S}(t+1) \rightarrow a(t+1)$.

Кратко опишем метод SARSA. В этом методе итеративно формируются оценки величины суммарной награды $Q(\mathbf{S}(t), a(t))$, которую получит анимат, если в ситуации $\mathbf{S}(t)$ он выполнит действие $a(t)$. Математическое ожидание награды равно:

$$Q(\mathbf{S}(t), a(t)) = \mathrm{E}\{r(t) + \gamma\, r(t+1) + \gamma^2\, r(t+2) + \ldots\} \mid \mathbf{S} = \mathbf{S}(t), a = a(t). \qquad (2)$$

Из (1) и (2) следует $Q(\mathbf{S}(t), a(t)) = E[r(t) + \gamma\, Q(\mathbf{S}(t+1), a(t+1))]$. Ошибку естественно определить так [15]:

$$\delta(t) = r(t) + \gamma\, Q(\mathbf{S}(t+1), a(t+1)) - Q(\mathbf{S}(t), a(t)). \qquad (3)$$

Величина $\delta(t)$ называется ошибкой временной разности.

Здесь $\delta(t)$ – разность между той оценкой суммарной величины награды, которая формируется у анимата для момента времени t после выбора действия $a(t+1)$ в следующей ситуации $\mathbf{S}(t+1)$ в момент времени $t+1$, и предыдущей оценкой этой же величины, которая была у анимата в момент времени t . Предыдущая оценка равна $Q(\mathbf{S}(t), a(t))$, новая оценка равна $r(t) + \gamma\, Q(\mathbf{S}(t+1), a(t+1))$, что и отражает формула (3) для величины $\delta(t)$. В соответствии с этим $\delta(t)$ анимат и обучается (см. ниже, формулу (4)).

Каждый такт времени происходит как выбор действия, так и обучение анимата. Выбор действия происходит так:

- в момент t с вероятностью 1 - ε выбирается действие с максимальным значением $Q(\mathbf{S}(t), a_j)$: $a(t) = \arg\max_j \{Q(\mathbf{S}(t), a_j)\}$;
- с вероятностью ε выбирается произвольное действие, $0 < \varepsilon << 1$.

Такую схему выбора действия называют «ε–жадным правилом».

Обучение, т.е. переоценка величин $Q(\mathbf{S}, a)$ происходит в соответствии с оценкой ошибки $\delta(t)$ – к величине $Q(\mathbf{S}(t), a(t))$ добавляется величина, пропорциональная ошибке временной разности $\delta(t)$:

$$\Delta Q(\mathbf{S}(t), a(t)) = \alpha\, \delta(t) = \alpha\, [r(t) + \gamma\, Q(\mathbf{S}(t+1), a(t+1)) - Q(\mathbf{S}(t), a(t))], \quad (4)$$

где α – параметр скорости обучения.

Метод временной разности идейно связан с методом динамического программирования, и в том и другом случае общая оптимизация многошагового процесса принятия решения происходит путем упорядоченной процедуры одношаговых оптимизирующих итераций, причем оценки эффективности тех или иных решений, соответствующие предыдущим шагам процесса, переоцениваются с учетом знаний о возможных будущих шагах. Например, при решении задачи поиска оптимального маршрута в лабиринте от стартовой точки к определенной целевой точке сначала находится конечный участок маршрута, непосредственно приводящий к цели, а затем ищутся пути, приводящие к конечному участку, и т.д. В результате постепенно прокладывается трасса маршрута от его конца к началу. Обучение с подкреплением, адаптивные критики и подобные методы часто называют приближенным динамическим программированием [39].

Важное достоинство метода обучения с подкреплением – его простота. Т.е. анимат получает от учителя или из внешней среды только сигналы подкрепления $r(t)$. Здесь учитель поступает с обучаемым объектом примитивно: "бьет кнутом" (если действия объекта ему не нравятся, $r(t) < 0$), либо "дает пряник" (в противоположном случае, $r(t) > 0$), не объясняя обучаемому объекту, как именно нужно действовать. Это радикально отличает этот метод от таких традиционных в теории нейронных сетей методов обучения, как метод обратного распространения ошибок, для которого учитель точно определяет, что должно быть на выходе нейронной сети при заданном входе.

Удалено: ах

Метод обучения с подкреплением был исследован рядом авторов (см. подробную библиографию в [15]) и был использован многочисленных приложениях. В частности, применения этого метода включают в себя:

- оптимизацию игры в триктрак (достигнут уровень мирового чемпиона);
- оптимизацию системы управления работы лифтов;
- формирование динамического распределения каналов для мобильных телефонов;
- оптимизацию расписания работ на производстве.

Подчеркнем, что метод обучения с подкреплением может рассматриваться как развитие автоматной теории адаптивного поведения, разработанной в работах М.Л. Цетлина и его последователей [4,5].

В свою очередь, метод обучения с подкреплением получил свое развитие в работах по адаптивным критикам, в которых рассматриваются нетривиальные методы обучения, использующие нейросетевые аппроксиматоры функций оценки качества функционирования анимата. Простейшие схемы адаптивных критиков приведем в следующем разделе. Более подробная характеристика адаптивных критиков содержится в статье Д.В. Прохорова в настоящей книге (глава 10).

2.4.2. Нейросетевые адаптивные критики [40]

Конструкции адаптивных критиков можно рассматривать как развитие моделей обучения с подкреплением на случай, когда как ситуации, так и действия задаются векторами **S** и **A** и изложенная выше схема итеративного формирования матрицы $Q(\mathbf{S}(t), a(t))$ не работает. В этом случае характеристики системы управления целесообразно представить с помощью параметрически задаваемых аппроксимирующих функций (например, с помощью искусственных нейронных сетей), а обучение проводить путем итеративной оптимизации параметров. В случае аппроксимации с помощью нейронных сетей, параметрами аппроксимирующих функций

являются веса синапсов нейросети, оптимизация производится путем подстройки весов, например, аналогично тому, как это делается в методе обратного распространения ошибки.

В конструкции аниматов на основе адаптивных критиков входят два важных блока системы управления: Критик и Контроллер (иногда используют также термин Актор).

Критик – это блок системы управления, который оценивает качество ее работы.

Контроллер – блок системы управления, формирующий действия этой системы.

Ниже мы опишем две простые конструкции адаптивных критиков: Q-критик и V-критик. Обе конструкции используют нейросетевую аппроксимацию характеристик системы управления.

Q-критик. Схема Q-критика представлена на рис. 4. Предполагаем, что как Критик, так и Контроллер представляют собой многослойные персептроны (т.е. нейронные сети, такие же, какие используются в методе обратного распространения ошибки) с весами синапсов $\mathbf{W_C}$ и $\mathbf{W_A}$, соответственно.

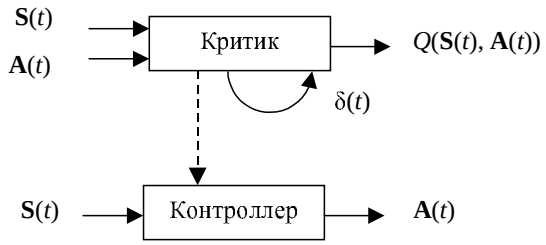


Рис. 4. Схема Q-критика.

Функционирование этой схемы происходит следующим образом. В момент времени t Контроллер по вектору входной ситуации $\mathbf{S}(t)$ определяет вектор действия $\mathbf{A}(t)$ (команды на эффекторы). На входы Критика подаются два вектора: $\mathbf{S}(t)$ и $\mathbf{A}(t)$. По этому составному вектору Критик делает оценку качества $Q(t) = Q(\mathbf{S}(t),\mathbf{A}(t))$ действия $\mathbf{A}(t)$ в текущей ситуации $\mathbf{S}(t)$. Действие $\mathbf{A}(t)$ выполняется, анимат получает награду $r(t)$. Далее происходит переход к следующему моменту времени t+1. Все операции повторяются, в том числе делается оценка значения $Q(t+1)$. После этого определяется ошибка временной разности:

$$\delta(t) = r(t) + \gamma\, Q\,(t+1) - Q\,(t). \qquad (5)$$

Обучение нейросетей выполняется следующим образом:

$$\Delta\, \mathbf{W_C} = \alpha_1\, \mathrm{grad}_{WC}(Q\,(t))\, \delta(t), \qquad (6)$$

$$\Delta\, \mathbf{W_A} = \alpha_2\, \Sigma_k\, \{[\partial Q\,(t)\, /\partial\, A_k(t)]\, \mathrm{grad}_{WA}\, A_k(t)\,\}, \qquad (7)$$

где α_1 и α_2 – параметры скорости обучения. Производные по весам синапсов $\mathrm{grad}_{WC}(.)$ и $\mathrm{grad}_{WA}(.)$ в (6) и (7), а также $\partial Q\,(t)\, /\partial\, A_k(t)$ в (7) рассчитываются как производные сложных функций, аналогично тому, как это делается в методе обратного распространения ошибки [26]. В формуле (7) учитывается, что нужно брать производные по всем компонентам вектора $\mathbf{A}(t)$ и суммировать по всем этим компонентам.

Смысл изменений весов синапсов по формулам (6),(7) состоит в том, что веса Критика и Контроллера меняются таким образом, чтобы уменьшить ошибку в оценке ожидаемой суммарной награды (обучение Критика) и увеличить значение самой награды при попадании анимата в близкие ситуации (обучение Контроллера).

V-критик. Схема V-критика, использующего модель, представлена на рис. 5.

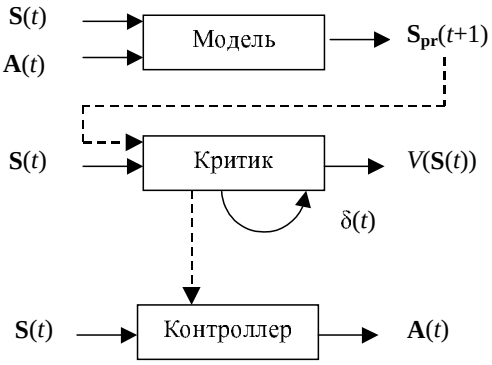


Рис. 5. Схема V-критика.

В этой схеме Критик, в отличие от схемы Q-критика, оценивает качество ситуации $V(\mathbf{S}(t))$ независимо от выполняемого действия. Однако такая схема управления содержит блок Модель, в котором прогнозируется будущее состояние $\mathbf{S_{pr}}(t+1) = \mathbf{S_{pr}}(\mathbf{S}(t), \mathbf{A}(t))$ в зависимости от текущего состояния $\mathbf{S}(t)$ и выполняемого действия $\mathbf{A}(t)$. И для этого прогнозируемого состояния $\mathbf{S_{pr}}(t+1)$ блок Критик может сделать оценку его качества $V_{pr} = V(\mathbf{S_{pr}}(t+1)) = V(\mathbf{S_{pr}}(\mathbf{S}(t), \mathbf{A}(t)))$.

Предполагаем, что Критик, Контроллер и Модель представляют собой многослойные персептроны с весами синапсов $\mathbf{W_C}$, $\mathbf{W_A}$ и $\mathbf{W_M}$, соответственно.

Функционирование этой схемы происходит следующим образом. В момент времени t Контроллер по вектору входной ситуации $\mathbf{S}(t)$ определяет вектор действия $\mathbf{A}(t)$. Критик делает оценку качества $V(t) = V(\mathbf{S}(t))$ текущей ситуации $\mathbf{S}(t)$. Модель прогнозирует следующее состояние $\mathbf{S_{pr}}(t+1) = \mathbf{S_{pr}}(\mathbf{S}(t), \mathbf{A}(t))$. Критик оценивает качество прогнозируемой ситуации $V_{pr} = V(\mathbf{S_{pr}}(t+1))$. Действие $\mathbf{A}(t)$ выполняется, анимат получает награду $r(t)$. Оценивается ошибка временной разности:

$$\delta(t) = r(t) + \gamma\, V(\mathbf{S_{pr}}(t+1)) - V(\mathbf{S}(t)) . \qquad (8)$$

Обучаются Критик:

$$\Delta\, \mathbf{W_C} = \alpha_1 \operatorname{grad}_{WC}(V(t))\, \delta(t), \qquad (9)$$

и Контроллер:

$$\Delta\, \mathbf{W_A} = \alpha_2 \Sigma_k \{[\partial\, V(\mathbf{S_{pr}}(t+1)/\, \partial\, A_k(t)] \operatorname{grad}_{WA} A_k(t)\} , \qquad (10)$$

$$\partial\, V(\mathbf{S_{pr}}(t+1)/\, \partial\, A_k(t) = \Sigma_j \{[\partial\, V/\, \partial S_{pr\,j}\,]\, [\partial S_{pr\,j}\, /\partial\, A_k(t)]\}. \qquad (11)$$

Производные в (11) берутся в соответствии с формулами нейронных сетей Критика и Модели.

Производится переход к следующему моменту времени $t+1$. Сравниваются прогнозируемая $\mathbf{S_{pr}}(t+1)$ и реальная ситуация $\mathbf{S}(t+1)$. В соответствии с ошибкой этого прогноза обучается Модель обычным методом обратного распространения ошибки.

Обучение Критика состоит в том, чтобы итеративно уточнять оценку качества ситуаций $V(\mathbf{S}(t))$ в соответствии с поступающими подкреплениями.

Обучение Контроллера состоит в том, чтобы постепенно формировать действия, приводящие к ситуациям с высокими значениями качества.

Смысл обучения Модели – уточнение прогнозов будущих ситуаций.

Отметим, что оценка функции качества $V(\mathbf{S}(t))$ в этой схеме, аналогична эмоциональной оценке текущего состояния системы в моделях А.А. Жданова (см. главу 13 в настоящей книге).

Более полно теория адаптивных критиков и ее современное состояние характеризуется в статье Д.В. Прохорова в настоящем сборнике (глава 10).

3. Исследования AnimatLab

Исследования по адаптивному поведению ведутся в ряде университетов и лабораторий. Кратко охарактеризуем работы одной из ведущих лабораторий – AnimatLab, которой руководит один из инициаторов этого направления исследований Жан-Аркадий Мейер.

3.1. Общий подход AnimatLab: Эволюция + обучение + онтогенез

Общий подход этой лаборатории можно охарактеризовать следующим образом [41]. Анимат (рис. 6) существует в реальной или модельной среде. У него есть сенсоры, которые воспринимают информацию из внешней и внутренней среды анимата, и эффекторы, посредством которых он взаимодействует со средой, а также система управления, которая координирует восприятие и действия анимата. Поведение анимата считается *адаптивным*, если система управления поддерживает жизненно важные переменные анимата (например, V_1 и V_2 на рис. 6) в допустимых пределах. На рис. 6 штриховая стрелка показывает возможную траекторию, выходящую за пределы допустимой области (серый фон – недопустимая область переменных). Сплошная стрелка показывает "исправленную" траекторию, откорректированную с помощью системы управления, обеспечивающей поддержание переменных в допустимой (светлой) области.

Если система управления выбирает последовательные цели, которые анимат стремится достичь, то о такой системе можно говорить как о *мотивационной системе* (motivational system). Система управления анимата может формироваться и модифицироваться путем *обучения*, индивидуального *развития* (онтогенеза) и *эволюции*.

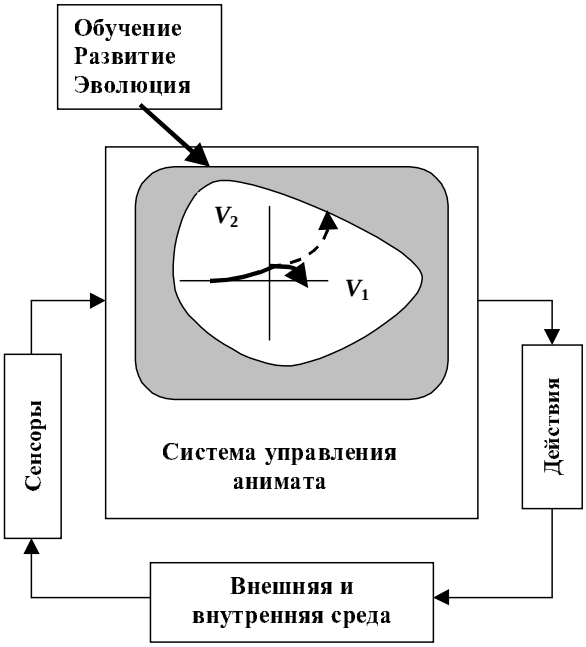


Рис. 6. Общая схема поведения анимата. Подход AnimatLab [41].

3.2. Онтогенез нейронной сети анимата – схема формирования структуры "нервной системы" робота

В работах AnimatLab была сделана интересная попытка промоделировать индивидуальное развитие, *"онтогенез" нейронной сети анимата* [42-44].

Общая схема метода состоит в следующем (рис. 7). Нейронная сеть анимата формируется с помощью специальной программы, контролирующей процесс конструирования сети. Эта программа имитирует процесс развития нейронной сети в процессе индивидуального взросления организма. Сама программа оптимизируется с помощью эволюционного алгоритма. Нейронная сеть формируется в двумерной ограниченной области. Программа состоит из инструкций (команд), которые определяют процессы возникновения новых нейронов (или исчезновения уже имеющихся нейронов) в этой области, формирование связей между нейронами и задание весов синаптических связей между нейронами. Инструкции программы составляют геном анимата. Работа формирующихся нейронных сетей оценивается по поведению анимата некоторой естественной функцией приспособленности, которая определяет отбор наиболее эффективных программ, кодируемых геномами аниматов. Так как структура сети определяется расположением нейронов в двумерной области, то метод называется геометрическим.

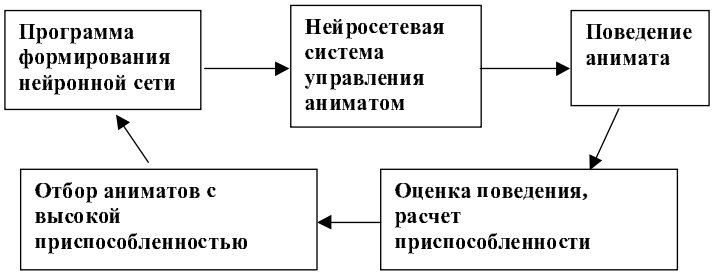


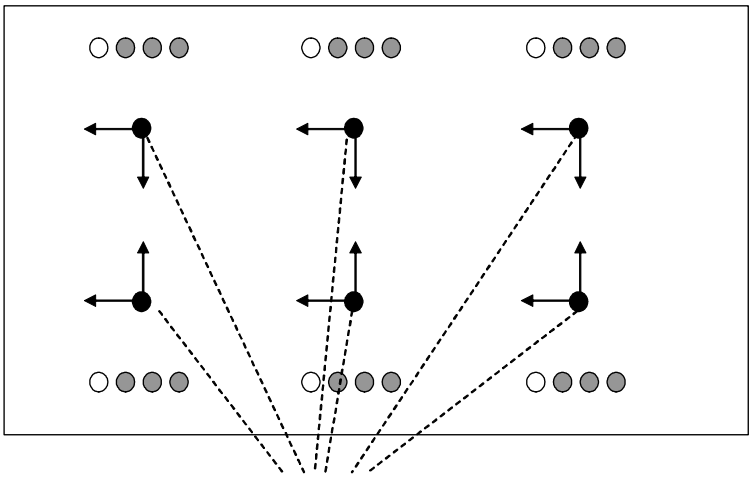
Рис. 7. Общая схема геометрического метода формирования структуры нейронной сети.

Эта схема иллюстрирует процесс формирования отдельного модуля нейронной сети. К такому модулю могут добавляться новые модули. Причем сначала формируются модули нижних уровней, определяющие инстинктивное, рефлекторное поведение анимата (например, согласованное движение 6-ти ног при прямолинейном перемещении анимата). А затем формируются модули более высоких уровней (например, модуль, управляющий остановкой и возобновлением прямолинейного движения), управляющие работой модулей нижних уровней.

Программа развития нейронной сети. Опишем принципы формирования нейронной сети на основе инструкций программы развития. Процесс формирования определяется как экспериментатором, конструирующим нейронную сеть, так и работой программ развития и эволюционной оптимизацией этих программ.

Первоначально экспериментатор выделяет двумерную область (например, прямоугольник). В этой области он задает расположение тех нейронов, которые заведомо необходимы (сенсорные и моторные нейроны). Затем задается множество затравочных нейронов, из которых будет развиваться нейронная сеть.

Пример области с расположенными на ней сенсорными, моторными и затравочными нейронами показан на рис. 8.



Программа развития нейронной сети

– сенсорные нейроны

– мотонейроны трех разных типов, управляющие разными "мышцами" анимата

– затравочные нейроны

Рис. 8. Схема расположения изначально задаваемых нейронов в прямоугольной двумерной области (по работе [42], с изменениями). Симметрия схемы отражает то, что она предназначена для формирования нейронной сети 6-ногого анимата. Каждый из затравочных нейронов имеет свою локальную систему координат, которую он затем использует при формировании дочерних нервных клеток и связей с другими нейронами.

Экспериментатор также задает инструкции, из которых будут формироваться программы развития нейросетей, и некоторые синтаксические ограничения на то, как из этих инструкций можно формировать программы.

В работах [42-44] была использована модель интегрирующего нейрона, согласно которой динамика мембранного потенциала i-й нервной клетки описывается уравнением:

$$\tau_i \cdot dm_i / dt = -m_i + \sum_j w_{ij} x_j + I_i ,$$

где $x_j = \{1+\exp [-(m_j + B_j)]\}^{-1}$ – частота импульсов нейрона, B_j – случайный порог нейрона со средним значением b_j , τ_i – постоянная времени релаксации i -го нейрона, I_i – внешний вход i -го нейрона от определенного сенсора, w_{ij} – синаптический вес, характеризующий связь от j -го к i -му нейрону.

Удалено: от

Примеры инструкций, определяющих программы развития нейронной сети, представлены в таблице 1.

Таблица 1. Инструкции программы развития нейронной сети

Удалено: – команды

DIVIDE (α, r)	Создать новый (дочерний) нейрон
GROW (α, r, w)	Создать соединение к другому нейрону
DRAW (α, r, w)	Создать соединение от другого нейрона
SETBIAS (b)	Изменить порог нейрона
SETTAU (τ)	Изменить постоянную времени релаксации нейрона
DIE	Удалить нейрон

Инструкции характеризуются параметрами, которые определяют, как именно будет развиваться нейронная сеть. Здесь α – угол, задающий направление к создаваемому нейрону или направление, в котором формируется новое соединение, r – длина соответствующего соединения, w – вес связи, b – величина порога и τ – время релаксации. Пример использования команд и параметров приведен ниже, при описании рис. 9.

Программа развития состоит из подпрограмм. Каждому затравочному нейрону соответствует своя подпрограмма.

Инструкции каждой из подпрограмм скомпонованы в граф, в котором есть корневой узел, команда-инструкция которого применяется к затравочному нейрону. Работа подпрограммы начинается с того, что затравочная нервная клетка исполняет инструкцию корневого узла подпрограммы.

Одна из инструкций (DIVIDE) соответствует делению клетки на материнскую и дочернюю. Формирование дочерней нервной клетки соответствует ветвлению графа подпрограммы; при этом после деления клетки (и соответствующего ветвления графа) инструкция левого узла относится к материнской нервной клетке, а инструкция правого узла – к дочерней клетке. Итак, инструкция DIVIDE соответствует узлу ветвления графа на две ветви. Все остальные инструкции не приводят к ветвлению графа подпрограммы.

Некоторые из команд-инструкций терминальные, после этих инструкций процесс развития той нервной клетки, к которой они применяются, останавливается.

Программы формирования нейронной сети оптимизируются с помощью эволюционного алгоритма. В процессе эволюции программы испытывают мутации и рекомбинации. Рекомбинации представляют собой обмен подграфами подпрограмм – аналогично тому, как это осуществляется в генетическом программировании [35,36]. Пример применения инструкций показан на рис. 9.

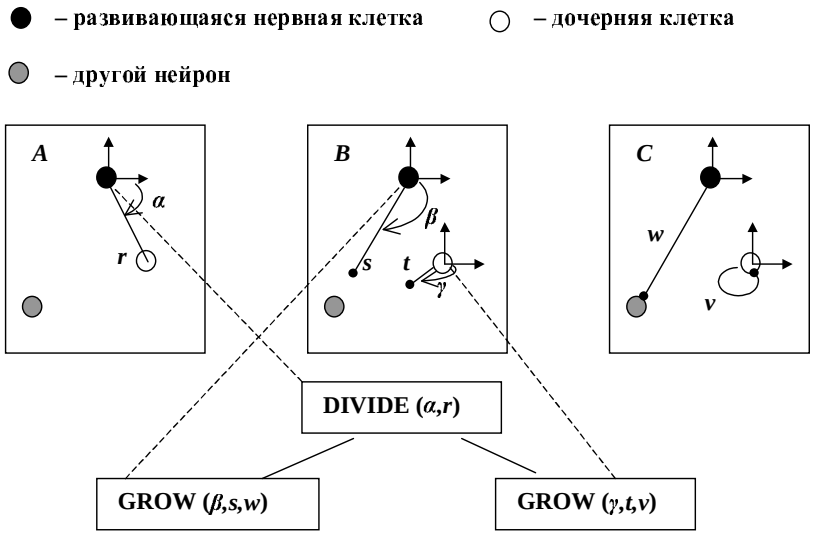


Рис. 9. Пример развития нейронной сети. Развитие происходит под управлением программы, представленной внизу рисунка. Программа содержит команду деления развивающейся нервной клетки DIVIDE (α,r) и команды формирования связей от материнской GROW (β,s,w) и дочерней GROW (γ,t,v) клеток. Первые два параметра в этих командах определяют углы и расстояния, в соответствии с которыми определяется положение дочерней клетки (для команды DIVIDE) и направление и длина синаптической связи (для команды GROW). Третий параметр команды GROW определяет величину синаптической связи. По [42] , с изменениями.

Как видно из рис. 9A, сначала по команде DIVIDE (α,r) формируется дочерняя клетка на расстоянии r от материнской с "азимутом" α . Дочерняя клетка наследует локальную систему координат материнской клетки. Затем (рис. 9B), как материнская (по команде GROW (β,s,w)) так и дочерняя (по команде GROW (γ,t,v)) клетки формируют отростки длинной s и t и под углом β и γ , соответственно. Далее (рис. 9C) концы этих отростков подсоединяются к *ближайшей из клеток* и задаются веса (w и v) соответствующих синаптических связей. На этом указанный блок программы выполнен.

Итак, геном анимата представляет собой определенную программу, однозначно определяющую процесс формирования структуры и весов нейронной сети. Программа состоит из инструкций, которые задают правила расстановки нейронов в заданной геометрической области и формирования синаптических связей между нейронами.

Эволюционный алгоритм. Эволюционный алгоритм в данном методе состоит в следующем. Поведение анимата оценивается в соответствии с эвристически задаваемой функцией приспособленности, производится отбор аниматов с большими приспособленностями, и наиболее приспособленные аниматы дают потомков.

При формировании потомков применяются три генетических оператора: 1) совместимая рекомбинация подграфов программ выбранных родителей, 2) формирование новых случайных (но допустимых) подграфов взамен старых, 3) и случайные мутации параметров, входящих в команды-инструкции.

Удалено: под-

Удалено: под-

В целом схема эволюции типична для генетического алгоритма, однако, необходимо соблюдение определенных условий для того, чтобы программы потомков не выходили за рамки ограничений, накладываемых на эти программы.

Принцип модульности. Примеры использования геометрического метода. Принцип модульности подразумевает, что модули нейронной сети формируются последовательно – сначала один модуль, затем следующий. Например, в работе [42] сначала был эволюционно сформирован 1-й модуль, определяющий согласованное движение 6-ти ног при прямолинейном одномерном перемещении анимата, а затем 2-й модуль, управляющий остановкой и возобновлением прямолинейного движения.

В работе [43] была сформирована нейронная сеть, управляющая 2-мерным движением 6-ногого анимата. При этом нейронная сеть состояла из трех модулей:
1-й модуль управлял движением анимата,
2-й модуль контролировал работу первого модуля и обеспечивал перемещение анимата к заданной цели (к источнику запаха),
3-й модуль обеспечивал перемещение в среде с препятствиями и был предназначен для минимизации столкновений анимата с препятствиями.

Модули формировались последовательно (сначала 1-й, потом 2-й, затем 3-й). При формировании 1-го модуля приспособленность анимата оценивалась по скорости его движения (чем больше скорость перемещения, тем выше приспособленность). При формировании 2-го и 3-го модулей 1-й модуль оставался неизменным, а формировались только связи от нейронов новых модулей к нейронам первого модуля. Приспособленность программ 2-го модуля оценивалась по способности анимата находить источник запаха. Приспособленность программ 3-го модуля оценивалась по способности анимата избегать препятствия, случайно разбросанные в области его движения. Программы формирования каждого из модулей были различными, более того, несколько различался синтаксис инструкций и графов, на основе которых формировались программы.

В полученной нейронной сети 1-й модуль содержал 38 интернейронов (сформированных нейронов, обеспечивающих связи между сенсорными нейронами и мотонейронами) и 100 межнейронных соединений; 2-й модуль содержал 6 интернейронов и 22 межнейронных соединения; 3-й модуль содержал 2 интернейрона и 6 межнейронных соединений.

В работах [42, 43] геометрический метод "выращивания" структуры нейронной сети был применен к задаче формирования системы управления аниматами, моделируемыми компьютерной программой. Т.е. аниматы "жили" только в компьютере. В работе [44] этот метод был применен к реальному 6-ногому роботу SECT, который был обучен перемещаться по двумерной плоскости и избегать столкновения с препятствиями. Было продемонстрировано, что имитация поведения робота в компьютерных программах согласуется с поведением реального робота.

Анализ геометрического метода формирования структуры нейронной сети показывает, что этот метод достаточно универсален и может быть применен к широкому классу систем управления адаптивным поведением. Основные принципы этого метода сводятся к следующему.

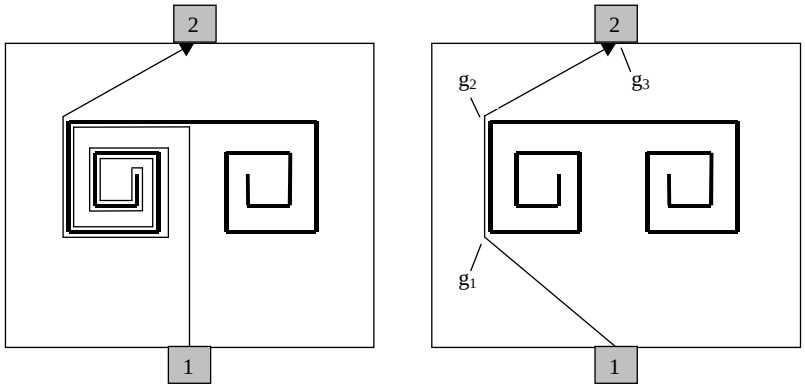
1. Создается специальный язык команд-инструкций, на основе которого строятся программы развития (онтогенеза) структуры и параметров нейронной сети.
2. Программы онтогенеза нейронных сетей оптимизируются эволюционным путем. Схема эволюционной оптимизации близка к таковой в генетическом программировании.
3. Сложная нейронная сеть строится по модульному принципу: сначала формируются модули нижних уровней управления, а затем модули верхних уровней иерархии управления.

3.3. Анимат MonaLysa – пример мотивационной системы

Интересное направление исследований AnimatLab – конструирование и моделирование мотивационных систем управления аниматами.

Пример мотивационной системы – довольно интеллектуальная архитектура управления аниматом MonaLysa, который, функционируя в сложной среде, способен сам выделять цели и подцели адаптивного поведения [12], MonaLysa – сокращение от MotivatiONAlLY autonomouS Animat. Основная идея данной системы управления состоит в том, что в процессе освоения внешнего мира и накопления опыта анимат стремится разбить задачу достижения глобальной цели на подзадачи, а затем использовать этот опыт при планировании решения новых задач.

В работе [12] исследовалось поведение анимата MonaLysa на примере навигационной задачи. Анимат помещался в центральную нижнюю точку прямоугольника, и нужно было попасть в центральную верхнюю точку (рис. 10), обходя различные препятствия. Анимат мог работать в "планирующем режиме", т.е., как сказано выше, разбивать задачи на подзадачи и планировать свои действия в соответствии с уже имеющимся опытом. Это поведение сравнивалось с поведением в "реактивном режиме" – без плана, на основе только текущей видимой ситуации.



а б

Рис. 10. Траектории движения анимата MonaLysa в реактивном (а) и планирующем (б) режиме работы системы управления. По [12] с изменениями. 1 – исходное положение анимата, 2 – конечная цель движения анимата. Жирной линией показаны препятствия, которые должен обойти анимат, тонкой линией – путь анимата.

В реактивном режиме анимат знает, где расположена конечная цель, и если нет препятствия, то движется прямо к этой цели; натыкаясь на препятствие, он обходит его до тех пор, пока не появится снова возможность двигаться к прямо к цели (рис. 10а). В планирующем режиме анимат на основании предшествующего опыта выделяет подцели (точки g_1, g_2, g_3 на рис. 10б), и движется прямо к текущей подцели, причем последняя подцель совпадает с конечной целью движения.

Отметим, что схема анимата MonaLysa была реализована как в компьютерной программе, так и для управления реальным роботом Khepera. Система управления в анимате MonaLysa была основана на простой версии классифицирующей системы (см. раздел 2.3).

Здесь мы охарактеризовали работы только одной лаборатории, ведущей исследования в области адаптивного поведения, в близких направлениях ведутся исследования и разработки в ряде университетских центров, таких как:

– Лаборатория искусственного интеллекта в университете Цюриха (руководитель Рольф Пфейфер) [45,46]. Основной подход этой лаборатории – познание природы интеллекта путем создания ("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) [47,48], ведущая исследования в области эволюционной роботики и принципов формирования адаптивного поведения.
– Лаборатория информатики и искусственного интеллекта в Массачусетском технологическом институте (руководитель Родни Брукс) [49,50], которая ведет исследования широкого спектра интеллектуальных и адаптивных систем, включая создание интеллектуальных роботов.

В нашей стране исследования адаптивного поведения пока ведутся скромными усилиями ученых-энтузиастов, некоторые работы которых представлены в настоящей книге.

4. Теория функциональных систем П.К. Анохина как концептуальная основа исследований адаптивного поведения

Для осмысления многообразия форм адаптивного поведения необходимо не только исследование конкретных моделей, но и разработка общих концепций и схем, позволяющих взглянуть сверху, "с высоты птичьего полета" на эти исследования.

Одной из таких концептуальных теорий может служить теория функциональных систем, предложенная и развитая в 1930-70 годах известным советским нейрофизиологом П.К. Анохиным [51-53].

Удалено: Из о

Функциональная система по П.К. Анохину – схема управления, нацеленного на достижение полезных для организма результатов.

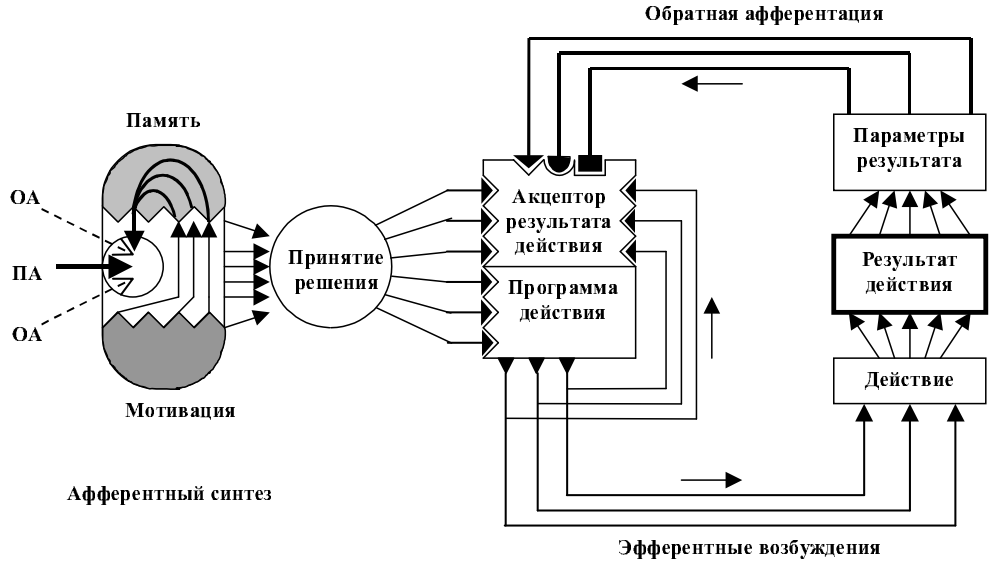


Рис. 11. Структура функциональной системы. ОА – обстановочная афферентация, ПА – пусковая афферентация.

Работа функциональной системы (рис. 11) может быть описана следующим образом.

Сначала происходит *афферентный синтез*, который включает в себя нейронные возбуждения, обусловленные 1) доминирующей мотивацией (понятие "мотивация" кратко обсуждается ниже), 2) обстановочной и пусковой афферентацией, 3) врожденной и приобретаемой памятью.

За афферентным синтезом следует *принятие решения*, при котором происходит уменьшение степеней свободы для эфферентного синтеза и выбор конкретного действия в соответствии с доминирующей мотивацией и с другими составляющими афферентного синтеза.

Затем следует формирование *акцептора результата действия*, т.е. прогноза результата. Прогноз включает в себя оценку параметров ожидаемого результата.

Эфферентный синтез – подготовка к выполнению действия. При эфферентном синтезе происходит генерация определенных нейронных возбуждений перед подачей команды на выполнение действия.

Все этапы достижения результата сопровождаются *обратной афферентацией*. Если параметры фактического результата отличаются от параметров акцептора результата действия, то действие прерывается и происходит новый афферентный синтез. В этом случае все операции повторяются, до тех пор, пока не будет достигнут конечный потребный результат.

Таким образом, функциональная система имеет циклическую (с обратными афферентными связями) саморегулирующуюся архитектонику.

Теория П.К. Анохина подразумевает *динамизм функциональных систем*. Для каждого конкретного поведенческого акта может быть сформирована своя функциональная система.

Функциональные системы формируются в процессе *системогенеза*. Теория системогенеза, которая исследует закономерности формирования функциональных систем в *эволюции*, *индивидуальном развитии* и *обучении* [54], может рассматриваться как отдельная ветвь теории функциональных систем. Отметим, что указанные составляющие системогенеза соответствуют составляющим формирования систем адаптивного поведения в трактовке AinimatLab (рис. 6).

Каждая функциональная система ориентирована на достижение *конечного потребного результата*.

Необходимо подчеркнуть, что теория функциональных систем была разработана, в первую очередь, для интерпретации нейробиологических данных и зачастую сформулирована в очень интуитивных терминах. Поэтому, хотя она и хорошо известна, она не общепризнана и практически не использовалась при разработке серьезных моделей адаптивного поведения. Можно сказать, что попытки формализации теории функциональных систем только начинаются [55-57]. Тем не менее, эта теория базируется на многочисленных биологических экспериментальных данных и представляет собой хорошую концептуальную основу для исследования широкого спектра проблем адаптивного поведения.

Отталкиваясь от теории П.К. Анохина, можно предложить общую кибернетическую схему управления целенаправленным адаптивным поведением естественного или искусственного организма (рис. 12). Здесь под организмом можно подразумевать как животное, так и робот или социально-экономическую систему: промышленную фирму, государство, человечество.

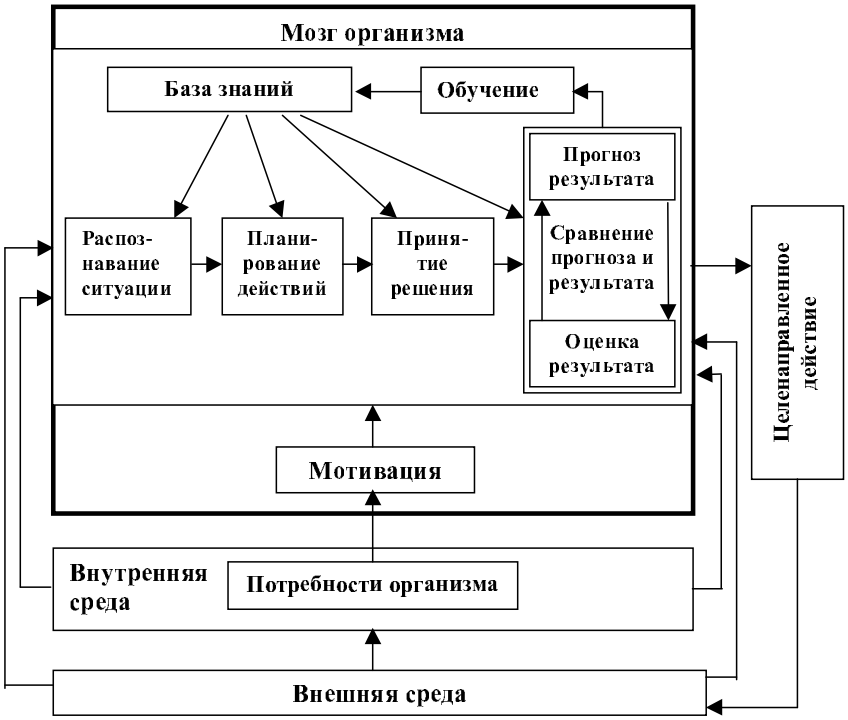


Рис. 12. Общая схема управления целенаправленным адаптивным поведением (в духе П.К. Анохина).

Что же можно делать сразу сейчас? Какие интересные задачи можно исследовать, отталкиваясь от теории функциональных систем?

Важное понятие функциональной системы – *мотивация*. Роль мотивации состоит в формировании цели и поддержке целенаправленных форм поведения. Мотивация может рассматриваться как активная движущая сила, которая стимулирует нахождение такого решения, которое адекватно потребностям организма в рассматриваемой ситуации. И имеет смысл провести моделирование эволюционного возникновения *целенаправленного* адаптивного поведения и анализ роли мотиваций в формировании целенаправленного поведения. Также следует отметить, что целенаправленность могла возникнуть на очень ранних стадиях эволюции, до появления каких-либо форм индивидуально приобретаемой памяти [58], поэтому, следуя пути, пройденному эволюцией, разумно начать с анализа этого свойства. Кроме того, свойство целенаправленности важно само по себе – это существенная особенность поведения *именно живых существ*.

Модели эволюционного возникновения целенаправленного адаптивного поведения были построены и исследованы в работах [59-61]. Основные результаты этого моделирования излагаются в следующем разделе.

5. Модели эволюционного возникновения целенаправленного адаптивного поведения

5.1. Модель «Кузнечик». Роль мотиваций в формировании адаптивного поведения [59,60]

В данной модели исследовался возможный механизм эволюционного возникновения целенаправленного поведения, обусловленного мотивациями.

Основные предположения модели состоят в следующем:

– Имеется популяция агентов (искусственных организмов), имеющих естественные потребности: 1) *потребность энергии* и 2) *потребность размножения*.

– Популяция эволюционирует в одномерной клеточной среде (рис. 13), в клетках может эпизодически вырастать трава (пища агентов). Каждый агент имеет *внутренний энергетический ресурс* R , который пополняется при съедании травы и расходуется при выполнении каких-либо действий. Уменьшение ресурса до нуля приводит к смерти агента. Агенты могут скрещиваться, рождая новых агентов.

– Потребности характеризуется количественно *мотивациями*. Если энергетический ресурс R агента уменьшается, то возрастает мотивация к пополнению энергетического ресурса (соответствующая потребности энергии) и уменьшается мотивация к размножению. При увеличении R мотивация к пополнению ресурса уменьшается, а мотивация к размножению растет.

– Поведение агента управляется его *нейронной сетью*. Сеть имеет один слой нейронов. На входы нейронов подаются сигналы, характеризующие внешнюю и внутреннюю среду агента, выходы нейронов определяют действия агента. Каждому возможному действию соответствует ровно один нейрон. В каждый такт времени совершается действие, соответствующее максимальному сигналу на выходе нейрона.

– Агенты "близорукие" – агент воспринимает состояние внешней среды только из трех клеток его поля зрения (рис. 13): той клетки, в которой агент находится, и двух соседних клеток.

– Агент может выполнять следующие *действия*: 1) быть в состоянии покоя ("отдыхать"), 2) двигаться, т.е. перемещаться на одну клетку вправо или влево, 3) прыгать через несколько клеток в случайную сторону, 4) есть (питаться), 5) скрещиваться. В силу способности агентов прыгать, мы называем их «кузнечиками».

– Нейронная сеть имеет специальные входы от мотиваций. Если имеется определенная мотивация, то поведение агента может меняться с тем, чтобы удовлетворить соответствующую потребность. Такое поведение можно рассматривать как *целенаправленное* (есть цель удовлетворить определенную потребность).

– Популяция агентов *эволюционирует*. Веса синапсов нейронной сети, управляющей поведением агента, составляют геном агента. Геном потомка формируется на основе геномов родителей при помощи рекомбинаций и мутаций.

– Мотивация к пополнению энергетического ресурса M_E и мотивация к размножению M_R определялись как простые функции энергетического ресурса агента R :

$$M_E = \max\{(R_0-R)/R_0\ ,\ 0\},\ \ M_R = \min\{R/R_1,\ 1\}\ ,$$

где R_0 , R_1 – параметры (обычно полагалось $R_0 = 2\ R_1$).

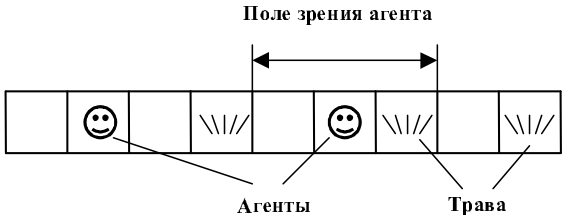


Рис. 13. Агенты в одномерной клеточной среде.

Модель исследовалась путем компьютерного моделирования эволюции популяции агентов. Нейронная сеть агентов исходной популяции определяла некоторые простые изначальные

инстинкты, обеспечивающие питание и размножение агентов. Далее наблюдалось, как в процессе эволюции изменялись нейронная сеть агентов и определяемое ей поведение агентов.

Для того чтобы исследовать влияние мотиваций на поведение агентов, были проведены две серии компьютерных экспериментов. В первой серии моделировалась эволюция популяции агентов с "выключенными" мотивациями (входы нейронов от мотиваций были "задавлены"), во второй серии мотивации "работали" (так, как это изложено выше).

Основные результаты проведенного моделирования таковы:

– Мотивации играют важную роль в исследованных эволюционных процессах. А именно, если сравнить популяцию агентов без мотиваций с популяцией агентов с мотивациями, то, как показывают компьютерные эксперименты, эволюционный процесс приводит к тому, что вторая популяция (с мотивациями) имеет значительные селективные преимущества по сравнению с первой (без мотиваций). Этот вывод иллюстрируется рис. 14.

– Анализ нейронных сетей и поведения агентов демонстрирует, что управление поведением агента без мотиваций (рис. 15) можно рассматривать как набор простых инстинктов (несколько отличающихся от изначально заданных), а управление агентом с мотивациями (рис. 16) – как *иерархическую систему управления*, состоящую из двух уровней: уровня простых инстинктов и метауровня, обусловленного мотивациями. При этом иерархическая система управления обеспечивает более эффективное управление, чем одноуровневая система, в которой поведение определяется одними лишь простыми инстинктами. Переход от схемы управления без мотиваций (рис. 15) к схеме управления с мотивациями (рис. 16) подобен метасистемному переходу от простых рефлексов к сложному рефлексу в теории метасистемных переходов В.Ф. Турчина [62].

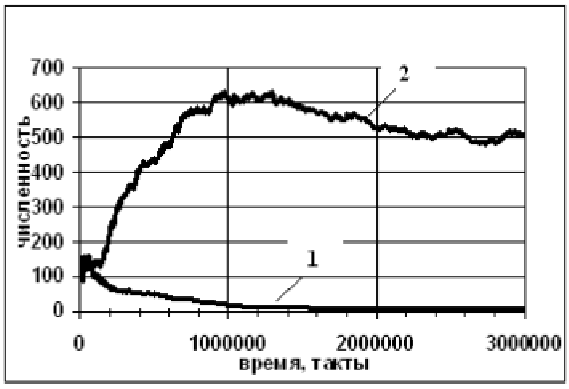


Рис. 14. Пример зависимостей численности популяции от времени для агентов без мотиваций (1) и с мотивациями (2). Видно, что популяция агентов с мотивациями имеет значительные селективные преимущества по сравнению с популяцией агентов без мотиваций.

Рис. 15. Схема управления агента без мотиваций. Поведение агента состоит из простых безусловных рефлексов, при котором выбор действия напрямую определяется текущим состоянием окружающей среды.

Рис. 16. Схема управления агента, обладающего мотивациями. Мотивации формируют новый уровень иерархии в системе управления агентами.

5.2. Возникновение естественной разветвленной иерархии целей [61]

Изложенная модель была развита в работе М.С. Бурцева [61], в которой исследовалось поведение популяции агентов в двумерном мире (рис. 17). При этом дополнительно в модель были введены 1) возможность борьбы между агентами и 2) эволюционное изменение структуры

нейронной сети, состоящей из рецепторов, эффекторов и связей между рецепторами и эффекторами.

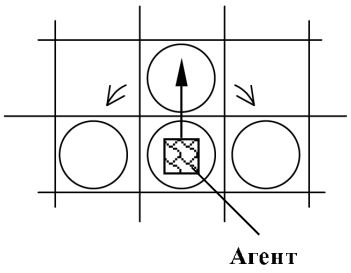


Рис. 17. Агент в двумерной клеточной среде. Агент ориентирован (стрелка показывает направление вперед), кружки – поле зрения агента. Действия агента: двигаться вперед, поворачиваться направо или налево, есть, размножаться, бороться с другими агентами. Система управления агента – однослойная нейронная сеть, оптимизируемая эволюционным методом.

Как и в предыдущей модели, для агентов исходной популяции задавалась некоторая минимальная система управления, обеспечивающая питание и размножение агентов. Поведение агентов начальной популяции (имеющих минимальный набор рецепторов и эффекторов) схематично представлено на рис. 18.

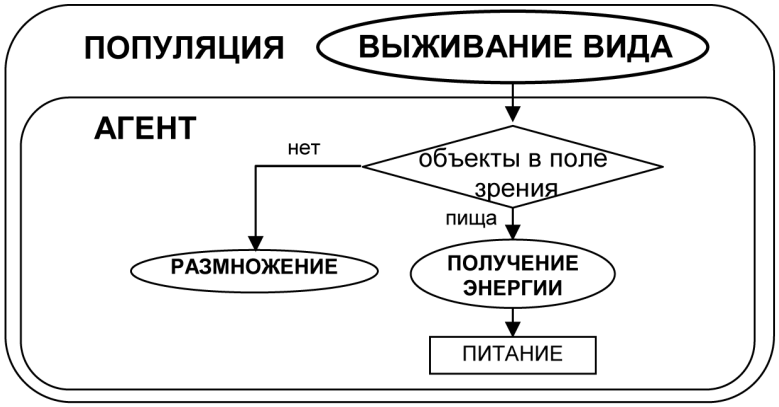


Рис. 18. Дерево условий для управления выбором подцелей агента начальной популяции.

В ходе эволюции поведение агентов структурируется. Стратегия агентов, сформированная в процессе эволюции, может быть представлена в виде схемы, показанной на рис. 19. Видно, что развивается достаточно сложное поведение, которое можно считать целенаправленным. Так первоначальный "инстинкт" агента, направленный на получение энергии, оптимизируется за счет появления еще одного уровня подцелей, направленных, соответственно: на собственно питание, на поиск пищи, на борьбу.

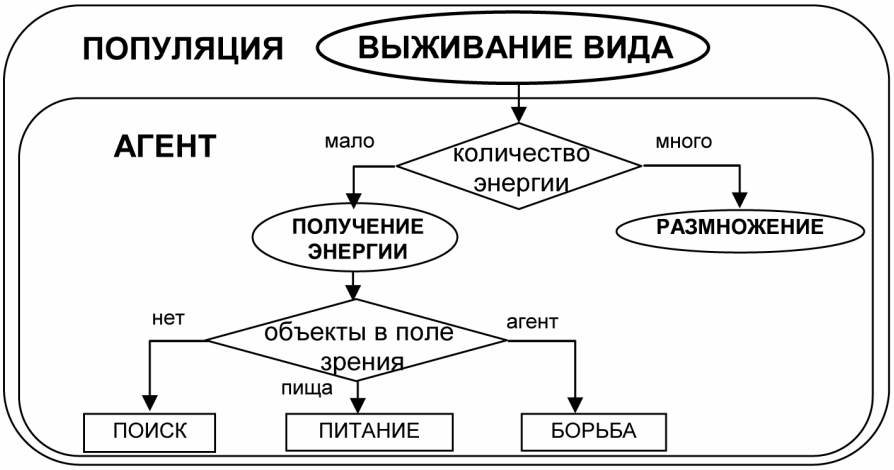


Рис. 19. Дерево условий для управления выбором подцелей, формирующееся в результате эволюции. Количество энергии здесь эквивалентно внутреннему энергетическому ресурсу агента *R*.

В целом моделирование, выполненное в работе [61], продемонстрировало, что в процессе исследованных эволюционных процессов возникает естественная иерархическая структура целей и подцелей.

6. Проект «Мозг Анимата» [63][1]

Очерченные модели пока еще очень фрагментарны и иллюстрируют только отдельные стороны адаптивного поведения. Поэтому целесообразно предложить общую «платформу» для систематического построения моделей адаптивного поведения. В работах [57, 63] предложен проект «Мозг Анимата», который как раз и нацелен на формирование общей схемы построения таких моделей. Кратко опишем данный проект.

Проект основан на теории функциональных систем П.К. Анохина [51-53] (см. раздел 4). В работе [57] была предложена первая версия архитектуры системы управления на основе нейросетевых блоков прогноза, обучаемых с помощью метода обратного распространения ошибки. Здесь мы опишем следующую версию архитектуры [63], основанную на нейросетевых адаптивных критиках (см. раздел 2.4.2).

Предполагается, что система управления аниматом имеет иерархическую архитектуру (рис. 20). Базовым элементом системы управления является отдельная функциональная система (ФС).

[1] Термин «Мозг Анимата» был предложен К.В. Анохиным.

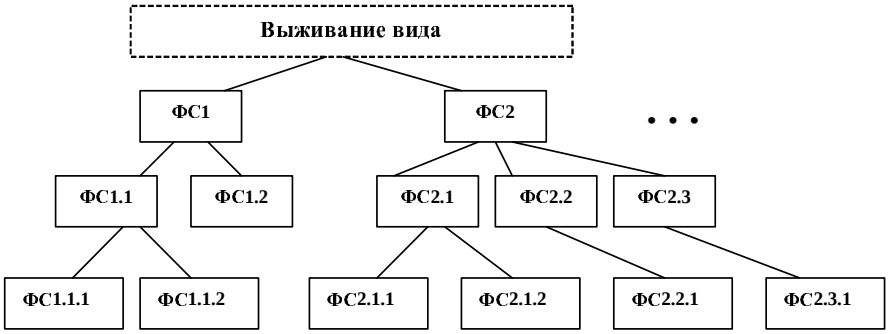


Рис. 20. Архитектура системы управления аниматом. ФС* – функциональная система.

Первый уровень (ФС1, ФС2, …) соответствует основным потребностям организма: питания, размножения, безопасности, накопления знаний. Более низкие уровни системы управления соответствуют тактическим целям поведения. Блоки всех этих уровней (включая первый) реализуются с помощью функциональных систем. Управление с верхних уровней может передаваться на нижние уровни (от «суперсистем» к «субсистемам») и возвращаться назад.

Самый верхний уровень соответствует выживанию вида (см. также схему иерархии управления на рис. 19). Этот уровень подразумеваемый, он не реализуется с помощью конкретной функциональной системы.

Предполагается, что система управления аниматом функционирует в дискретном времени и каждый такт времени активна только одна ФС.

Рассматривается простая формализация функциональной системы на основе нейросетевых адаптивных критиков, которая моделирует следующие важные особенности биологического прототипа: 1) принятие решения, 2) прогноз результата действия, 3) сравнение прогноза и результата и 4) коррекцию прогноза путем обучения в соответствующих нейронных сетях.

Функциональная система использует одну из возможных схем адаптивных критиков, представленную ниже.

6.1. Схема адаптивного критика

Рассматриваемая схема адаптивного критика состоит из двух блоков: Модель и Критик (рис. 21). Предполагается, что Модель и Критик – многослойные нейронные сети, и что производные по весам синапсов этих нейронных сетей могут быть вычислены обычным методом обратного распространения ошибки [26]. Адаптивный критик предназначен для выбора одного из нескольких действий. Например, при управлении движением действиями могут быть: двигаться вперед, поворачивать вправо, поворачивать влево, стоять на месте. В каждый момент времени t адаптивный критик должен выбрать одно из таких действий.

Цель адаптивного критика – максимизировать функцию суммарной награды $U(t)$:

$$U(t) = \sum_{j=0}^{\infty} \gamma^j r(t_j), \quad t = t_0, t_1, t_2, \ldots, \tag{12}$$

где $r(t_j)$ – текущее подкрепление (награда, $r(t_j) > 0$ или наказание, $r(t_j) < 0$), полученное адаптивным критиком в данный момент времени t_j, γ – коэффициент забывания, $0 < \gamma < 1$. Если не оговорено противное, предполагается, что $\tau = t_{j+1} - t_j = \text{const}$.

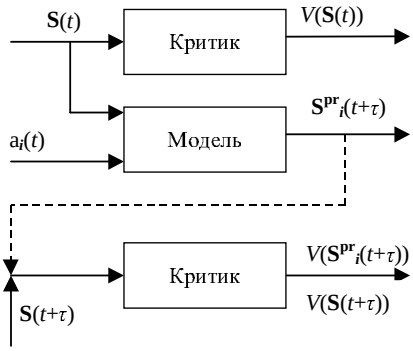


Рис. 21. Схема адаптивного критика, используемая в функциональной системе. Модель предсказывает следующую ситуацию $\mathbf{S}^{pr}_i(t+\tau)$ для всех возможных действий a_i, i=1,2,…, n_a. Текущая ситуация $\mathbf{S}(t)$, прогнозы $\mathbf{S}^{pr}_i(t+\tau)$ и реальная следующая ситуация $\mathbf{S}(t+\tau)$ подаются на вход Критика (одна и та же нейронная сеть Критика показана в два последовательных момента времени), на выходе которого формируются оценки качества ситуаций $V(\mathbf{S}(t))$, $V(\mathbf{S}^{pr}_i(t+\tau))$ и $V(\mathbf{S}(t+\tau))$.

Модель имеет два типа входов: 1) входы, характеризующие текущую ситуацию $\mathbf{S}(t)$ (сигналы из внешней и внутренней среды анимата), и 2) входы, характеризующие действия. Предполагается, что каждое возможное действие a_i кодируется своей собственной комбинацией входов, и что число возможных действий невелико. Роль Модели – прогноз следующей ситуации для всех возможных действий a_i, i=1,2,…, n_a.

Роль Критика – оценка качества ситуации $V(\mathbf{S})$ для текущей ситуации $\mathbf{S}(t)$, следующей ситуации $\mathbf{S}(t+\tau)$ и прогнозируемых ситуаций $\mathbf{S}^{pr}_i(t+\tau)$ для всех возможных действий.

В каждый момент времени выполняются следующие операции:

1) Модель предсказывает следующую ситуацию $\mathbf{S}^{pr}_i(t+\tau)$ для всех возможных действий a_i, i=1,2,…, n_a .

2) Критик оценивает качество ситуации для текущей ситуации $V(t) = V(\mathbf{S}(t))$ и всех прогнозируемых ситуаций $V^{pr}_i(t+\tau) = V(\mathbf{S}^{pr}_i(t+\tau))$. Величины V – оценки функции суммарной награды $U(t)$.

1) Применяется ε -жадное правило [15], а именно, выбирается действие следующим образом:

- с вероятностью 1 - ε выбирается действие с максимальным значением $V(\mathbf{S}^{pr}_i(t+\tau))$:
$k = \arg\max_i\{V(\mathbf{S}^{pr}_i(t+\tau))\}$
- с вероятностью ε выбирается произвольное действие a_k, $0 < \varepsilon << 1$,

где k – индекс выбираемого действия a_k .

4) Действие a_k выполняется.

5) Оценивается текущее подкрепление $r(t)$ и происходит переход к следующему моменту времени $t+\tau$. Наблюдается следующая ситуация $\mathbf{S}(t+\tau)$ и сравнивается с прогнозом $\mathbf{S}^{pr}_k(t+\tau)$. Корректируются веса $\mathbf{W}_M$ нейронной сети Модели с целью минимизации ошибки прогноза:

$$\Delta \mathbf{W}_M = \alpha_M \, \mathrm{grad}_{WM}(\mathbf{S}^{pr}{}_k(t+\tau))^T(\mathbf{S}(t+\tau) - \mathbf{S}^{pr}{}_k(t+\tau)) \ , \quad (13)$$

где α_M – скорость обучения нейронной сети Модели.

6) Критик оценивает величину $V(\mathbf{S}(t+\tau))$. Считается ошибка временной разности [15]:

$$\delta(t) = r(t) + \gamma \, V(\mathbf{S}(t+\tau)) - V(\mathbf{S}(t)) \ . \quad (14)$$

7) Корректируются веса $\mathbf{W}_C$ нейронной сети Критика:

$$\Delta \mathbf{W}_C = \alpha_C \, \delta(t) \, \mathrm{grad}_{WC}(V(t)) \ , \quad (15)$$

где α_C – скорость обучения нейронной сети Критика. Градиенты $\mathrm{grad}_{WM}(\mathbf{S}^{pr}{}_k(t+\tau))$ и $\mathrm{grad}_{WC}(V(t))$ означают производные выходов нейронных сетей относительно соответствующих весов синапсов. Градиенты считаются так же, как в методе обратного распространения ошибки.

Смысл обучения Модели – уточнение прогноза ожидаемых ситуаций.

Смысл обучения Критика – уточнение оценок качества ситуаций в соответствии с поступающими подкреплениями.

Изложенная схема адаптивного критика – ядро рассматриваемой функциональной системы. Множество функциональных систем формируют полную систему управления аниматом (рис. 20).

6.2. Функционирование системы управления аниматом

Структура ФС представлена на рис. 22. В основу ФС положена изложенная выше схема адаптивного критика. Дополнительные свойства ФС по сравнению со схемой адаптивного критика таковы: 1) ФС формирует команды субсистемам и посылает отчеты о результатах действий суперсистеме, и 2) сравнение между прогнозом $\mathbf{S}^{pr}{}_k(t+\tau)$ и результатом $\mathbf{S}(t+\tau)$ может быть отложено до момента $t+\tau$, когда поступит отчет от субсистем (детальней см. ниже). Связи данной ФС с супер/субсистемами показаны вертикальными жирными/пунктирными стрелками.

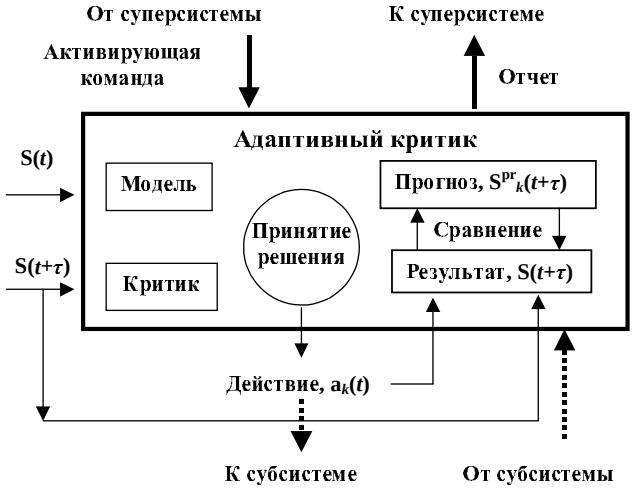


Рис. 22. Схема функциональной системы на основе адаптивного критика.

Работа ФС в рамках функционирования всей системы управления аниматом происходит следующим образом. ФС активизируются командой от суперсистемы. Модель и Критик функционируют так же, как описано выше в схеме работы адаптивного критика. В результате осуществляется выбор действия a_k. Дальнейшее зависит от вида действия a_k.

Если действие – команда для исполнительных элементов (сплошная стрелка вправо), то такое действие выполняется сразу; в этом случае $\tau = \tau_{min}$ – интервал между тактами времени минимален. Далее анимат получает подкрепление r из внешней или внутренней среды, и производится обучение в нейронных сетях адаптивного критика.

Другой тип действий – команды для субсистем (пунктирная стрелка вниз). Для такого действия подается команда активизации определенной субсистемы (выбор конкретной субсистемы определяется номером действия a_k). В этом случае сравнение прогноза и результата, оценка подкрепления r и обучение нейронных сетей откладывается до получения отчета от субсистемы, то есть до момента $t + \tau$, где $\tau > \tau_{min}$.

Обучение в обоих случаях осуществляется изложенным выше способом (раздел 6.1).

После выполнения всех этих действий, ФС посылает отчет об окончании своей работы соответствующей суперсистеме.

Описанный способ работы ФС представляет собой обычный режим функционирования. Вводится также экстраординарный режим, который имеет место, если прогноз существенно отличается от фактического результата: $\| \mathbf{S}^{pr}{}_k(t_j) - \mathbf{S}(t_j) \| > \Delta > 0$, где $\|.\|$ обозначает некоторую норму, например, евклидову. Предполагается, что в экстраординарном режиме величина ε (вероятность выбора случайного действия) в данной ФС и ее субсистемах резко возрастает, и поиск новых решений включает большую случайную компоненту. Этот поиск может сопровождаться случайным формированием и селекцией новых функциональных систем, аналогично селекции нейронных групп в теории нейродарвинизма Дж. Эдельмана [64]. Таким образом, обычный режим функционирования может рассматриваться как тонкая настройка системы управления аниматом, в то время как экстраординарный режим – это грубый поиск подходящего адаптивного поведения в чрезвычайных ситуациях.

Удалено: E

В данную схему управления поведением анимата несложно включить процедуру прерывания верхними уровнями работы нижних уровней иерархии функциональных систем, с помощью специальных связей между ФС. Например, если в ФС1, отвечающую за безопасность, поступил сигнал, характеризующий серьезную опасность для жизни анимата, а анимат занимался поиском "пищи" в дереве решений, "возглавляемом" ФС2, ответственной за потребность питания, то ФС1 имеет право прервать работу ФС2 и дать команду на избежание опасности.

Память о старых навыках в нейронных сетях ФС может «портиться» при обучении новым навыкам, что соответствует известной дилемме пластичности-стабильности. Рассматриваемая архитектура системы управления аниматом позволяет включить естественным образом долговременную память о приобретенных навыках. Если некоторый тип поведения был хорошо апробирован, то соответствующая ему ФС может быть скопирована в долговременную память, а именно, в ФС, в которой величины ε и α_C, α_M равны нулю. Обе ФС – долгосрочная и краткосрочная, с долговременной и кратковременной памятью, соответственно – могут играть одну и ту же роль в общей архитектуре системы управления. Для надежных навыков долгосрочная ФС имеет приоритет по отношению к краткосрочной. Однако если прогнозы ситуаций $\mathbf{S}^{pr}$, сделанные долгосрочной ФС, начинают отличаться от фактических $\mathbf{S}$, то управление возвращается к краткосрочной ФС.

Удалено: and

Итак, предложенная архитектура системы управления обеспечивает общий подход к моделированию адаптивным поведением анимата с естественными потребностями и соответствующими целями и подцелями. Сразу надо отметить, что использование адаптивных критиков в качестве функциональных систем – только один из возможных вариантов конструирования таких систем управления. Тем не менее, изложенная схема Мозга Анимата

позволяет уже сразу начинать работу по разработке конкретных моделей адаптивного поведения. Например, одной из первых модельных реализаций могло бы быть воспроизведение адаптивного поведения агентов с иерархией целей и подцелей, описанного в разделе 5.2 (см. рис. 19). Подчеркнем, однако, что роль проекта «Мозг Анимата» может быть и более серьезной: этот проект может быть положен в основу базовых моделей «интеллектуальных» изобретений биологической эволюции (см. рис. 23 ниже и следующий раздел) и анализа ступеней когнитивной эволюции.

7. Проблема происхождения интеллекта

Исследования адаптивного поведения можно связать с дальней стратегической задачей – задачей моделирования происхождения интеллекта человека. Философские и методологические основания к постановке исследований проблемы происхождения интеллекта обсуждаются в данном разделе.

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 $\{A, A \text{ --> } B\} \text{ => } B$.

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 --> БС, делает элементарный "вывод" {УС, УС --> БС } => 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 качественно аналогичны.

Мы можем пойти и дальше – можем представить, что в памяти животного есть *семантическая сеть*, сеть, узлами которой являются понятия, образы, а связи характеризуют взаимоотношения между понятиями. Можно далее представить процессы формирования разнообразных семантических сетей в процессе накопления жизненного опыта. Такие семантические сети, формируемые в памяти животных, по-видимому, аналогичны семантическим сетям, исследуемым разработчиками искусственного интеллекта [65].

Итак, мы можем думать над эволюционными корнями логики, мышления, интеллекта. И более того, было бы очень интересно попытаться строить 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 [66] (рис. 23). То есть,

целесообразно с помощью моделей представить общую картину эволюции когнитивных способностей животных и эволюционного происхождения интеллекта человека.

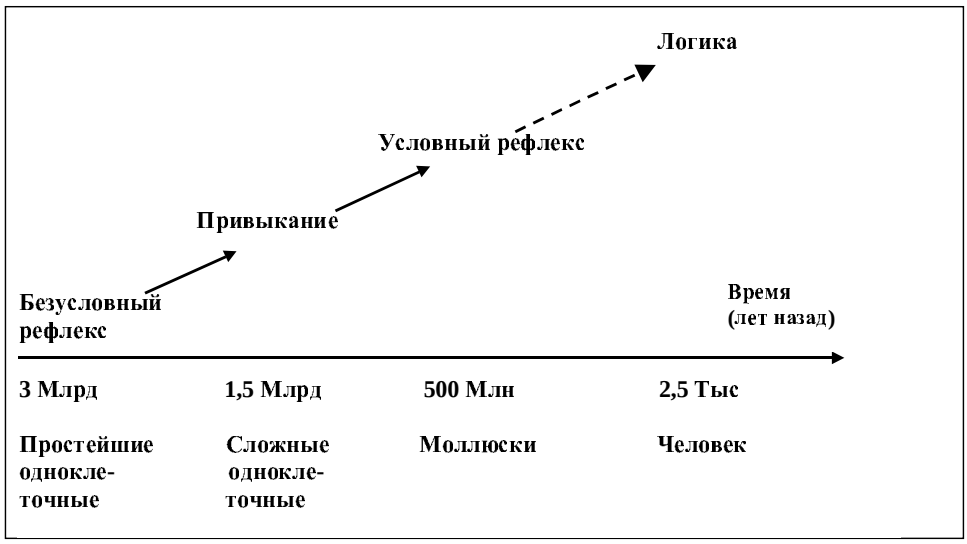


Рис. 23. "Интеллектуальные изобретения" биологической эволюции. "Авторы изобретений" и "даты приоритетов" представлены довольно условно.

Естественно, что такие исследования – это огромный фронт работы, и задачу построения теории происхождения интеллекта, задачу моделирования когнитивной эволюции можно пока рассматривать как сверхзадачу. Тем не менее, эта задача очень интересна и очень важна с точки зрения развития научного миропонимания. Исследования этой проблемы могли бы обеспечить определенное обоснование применимости нашего мышления в научном познании, то есть, укрепить фундамент всего величественного здания науки.

Подробнее задача моделирования "интеллектуальных изобретений" биологической эволюции обсуждается в работах [8,67,68].

Отметим, что задача моделирования когнитивной эволюции близка к сформулированной выше (раздел 1) программе-максимум исследований адаптивного поведения, и, естественно, что работы в этой области уже ведутся. Общее состояние моделей адаптивного поведения в контексте исследования когнитивной эволюции примерно таково. Есть множество математических и компьютерных моделей, характеризующих "интеллектуальные" изобретения: модель возникновения безусловного рефлекса на молекулярно-генетическом уровне [69], модели привыкания [70,71], большое количество моделей условных рефлексов [7, 72-76]. Однако эти модели очень фрагментарны, слабо разработаны и пока далеко не формируют общую картину эволюционного происхождения мышления, логики, интеллекта.

Отметим также потенциальное значение моделирования когнитивной эволюции для развития направления исследований Искусственный интеллект (ИИ) [77]. Это направление испытывает взлеты, падения, периоды энтузиазма и периоды разочарования. Скорее всего, это направление можно рассматривать как прикладное – применение принципов естественного интеллекта в искусственных практически важных для человека компьютерных системах. Судьба прикладных разработок зависит от наличия достаточно серьезного научного фундамента, на котором базируются такие разработки. Например, научной базой развития микроэлектроники во второй половине 20-го века была физика твердого тела. При этом для физиков чисто научные

исследования твердого тела были интересны практически 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 будут очень интересны для ученых. Но эти исследования могут быть тесно связаны и с разработками ИИ. И, следовательно, могло бы быть взаимное обогащение фундаментальных и прикладных исследований природы интеллекта. И, тем самым, исследования когнитивной эволюции могли бы служить научной основой разработок систем ИИ.

Автор благодарен Д.В. Прохорову за многочисленные консультации по моделям адаптивных критиков и за совместную работу над проектом «Мозг Анимата».

Литература:

1. 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.
2. Langton C. G., Taylor C., Farmer J. D., Rasmussen S. (Eds.) Artificial Life II: Proceedings of the Second Artificial Life Workshop. – Redwood City CA: Addison-Wesley, 1992.
3. Meyer J.-A., Wilson S.W. (Eds) From Animals to Animats. Proceedings of the First International Conference on Simulation of Adaptive Behavior. – The MIT Press: Cambridge, Massachusetts, London, England, 1990.
4. Цетлин М.Л. Исследования по теории автоматов и моделирование биологических систем. – М.: Наука, 1969. 316 с.
5. Варшавский В.И., Поспелов Д.А. Оркестр играет без дирижера. – М.: Наука, 1984.
6. Бонгард М.М., Лосев И.С., Смирнов М.С. Проект модели организации поведения – Животное // Моделирование обучения и поведения. – М.: Наука, 1975. С.152-171.
7. Гаазе-Рапопорт М.Г., Поспелов Д.А. От амебы до робота: модели поведения. – М.: Наука, 1987.
8. Редько В.Г. Эволюционная кибернетика. М.: Наука, 2001, 156 с.
9. Guillot A., Meyer J.-A. From SAB94 to SAB2000: What's new, Animat? // In Meyer et al. (Eds). From Animals to Animats 6. Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior. The MIT Press. 2000. See also: http://animatlab.lip6.fr/index.en.html
10. Непомнящих В.А. Аниматы как модель поведения животных // IV Всероссийская научно-техническая конференция "Нейроинформатика-2002". Материалы дискуссии "Проблемы интеллектуального управления – общесистемные, эволюционные и нейросетевые аспекты". М.: МИФИ, 2003. С. 58-76. См. также http://www.keldysh.ru/pages/BioCyber/RT/Nepomn.htm
11. Непомнящих В.А. Поиск общих принципов адаптивного поведения живых организмов и аниматов // Новости искусственного интеллекта. 2002. N. 2. С. 48-53.
12. Donnart 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. See also: http://animatlab.lip6.fr/index.en.html
13. 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).
14. Holland J.H., Holyoak K.J., Nisbett R.E., Thagard P. Induction: Processes of Inference, Learning, and Discovery. – Cambridge, MA: MIT Press, 1986.
15. Sutton R., Barto A. Reinforcement Learning: An Introduction. – Cambridge: MIT Press, 1998. See also: http://www.cs.ualberta.ca/~sutton/book/the-book.html
16. Потапов А.Б., Али М.К. Нелинейная динамика обработки информации в нейронных сетях // Новое в синергетике: Взгляд в третье тысячелетие. (Под ред. Г.Г. Малинецкого и С.П. Курдюмова). М.: Наука, 2002. С. 367-426.

17. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. М.: Эдиториал УРСС, 2002. 352 с.
18. Мак-Каллок У.С., Питтс У. Логическое исчисление идей, относящихся к нервной активности // Автоматы, под ред. Шеннона К.Э. и Маккарти Дж. М.: ИЛ, 1956. С. 362 - 384.
19. Фон Нейман Дж. Теория самовоспроизводящихся автоматов. М.: Мир, 1971, 382 с.
20. Фон Нейман Дж. Вероятностная логика и синтез надежных организмов из ненадежных компонент // Автоматы, под ред. Шеннона К.Э. и Маккарти Дж. М.: ИЛ, 1956. С. 68 - 139.
21. Розенблат Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга. М.: Мир, 1965.
22. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities // Proc. Natl. Acad. Sci. USA. 1982. V.79. N.8. PP.2554-2558.
23. Hopfield J.J. Neurons with gradual response have collective computational properties like those of two-state neurons // Proc. Natl. Acad. Sci. USA. 1984. V.81. N.10. PP.3088-3092.
24. Фролов А.А., Муравьев И.П. Нейронные модели ассоциативной памяти. М.: Наука, 1987. 160 с.
25. Фролов А.А., Муравьев И.П. Информационные характеристики нейронных сетей. М.: Наука, 1988. 160 с.
26. Rumelhart D.E., Hinton G.E., Williams R.G. Learning representation by back-propagating error // Nature. 1986. V.323. N.6088. PP. 533-536.
27. Carpenter G.A., Grossberg S. A massively parallel architecture for selforganizing neural pattern recognition machine // Comput. Vision, Graphics, Image Process. 1987. V.37. N.1. PP. 54-115.
28. Kohonen T. Self-organized formation of topologically correct feature maps // Biol. Cybern. 1982. V.43. N.1. PP. 56-69.
29. Fukushima K. Neocognitron: A hierarchical neural network capable for visual pattern recognition // Neural Networks. 1988. V.1. N.2. PP. 119-130.
30. Hopfield J.J., Tank D.W. Computing with neural circuits: A model // Science. 1986. V.233. N.464. PP. 625-633.
31. Hebb D.O. The organization of behavior. A neuropsychlogical theory. N.Y.: Wiley & Sons, 1949. 355 p.
32. Goldberg D.E. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, 1989.
33. Mitchell M. An Introduction to Genetic Algorithms. MIT Press, Cambridge, MA, 1996.
34. Курейчик В.М. Генетические алгоритмы и их применение. – Таганрог, ТРТУ, 2002.
35. Koza J. Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press, 1992.
36. Koza J. Genetic Programming II: Automatic Discovery of Reusable Subprograms. The MIT Press, 1994.
37. Holland J. H., Booker L.B., Colombetti M., Dorigo M., Forrest S., Goldberg D. G., Riolo R. L., Smith R. E., Lanzi P. L., Stolzmann W., Wilson S. W. What is a learning classifier system? // Holland J. H. et. al. (Eds). Learning Classifier Systems. Springer Verlag, 2000. pp. 3-32. See also: http://www.cs.unm.edu/~forrest/gacs_papers.htm
38. Klopf A. H. The Hedonistic Neuron: A Theory of Memory, Learning, and Intelligence. Hemisphere, Washington, 1982. 140 p.
39. Learning and Approximate Dynamic Programming: Scaling Up to the Real World (Edited by Jennie Si, Andrew Barto, Warren Powell, and Donald Wunsch), IEEE Press and John Wiley & Sons, 2004.
40. 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.
41. Сайт AnimatLab: http://animatlab.lip6.fr/index.en.html

42. Kodjabachian J., Meyer J.-A. Evolution and development of modular control architectures for 1-D locomotion in six-legged animats // Connection Science. 1998. V. 10. PP. 211-237. See also: http://animatlab.lip6.fr/index.en.html
43. Kodjabachian J., Meyer J.-A. Evolution and development of neural controllers for locomotion, gradient-following, and obstacle-avoidance in artificial insects // IEEE Transactions on Neural Networks. 1998. Vol. 9. PP. 796-812. See also: http://animatlab.lip6.fr/index.en.html
44. Filliat D., Kodjabachian J., Meyer J.-A. Incremental evolution of neural controllers for navigation in a 6-legged robot // Sugisaka and Tanaka (Eds). Proceedings of the Fourth International Symposium on Artificial Life and Robotics. Oita Univ. Press, 1999. See also: http://animatlab.lip6.fr/index.en.html
45. Сайт AI Laboratory of Zurich University: http://www.ifi.unizh.ch/groups/ailab/
46. Pfeifer R., Scheier C. Understanding Intelligence. MIT Press, 1999.
47. Сайт Laboratory of Artificial Life and Robotics: http://gral.ip.rm.cnr.it/
48. Nolfi S., Floreano D. Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines. Cambridge, MA: MIT Press/Bradford Books, 2000. 384 p.
49. Сайт MIT Computer Science and Artificial Intelligence Laboratory: http://www.csail.mit.edu/index.php
50. Brooks R.A. Cambrian Intelligence: The Early History of the New AI. MIT Press, 1999.
51. Анохин П.К. Принципиальные вопросы общей теории функциональных систем // Принципы системной организации функций. – М.: Наука, 1973. См. также: http://www.keldysh.ru/pages/BioCyber/RT/Functional.pdf
52. Анохин П.К. Системные механизмы высшей нервной деятельности. М.: Наука, 1979. 453 с.
53. Анохин П.К. Очерки по физиологии функциональных систем. – М.: Медицина, 1975.
54. Судаков К.В. (ред.). Теория системогенеза. – М.: Горизонт, 1997.
55. Умрюхин Е.А. Механизмы мозга: информационная модель и оптимизация обучения. М. 1999. 96 с.
56. Моделирование функциональных систем (под ред. Судакова К.В. и Викторова В.А.). – М.: РАМН, РСМАН, 2000. 254 с.
57. Анохин К.В., Бурцев М.С., Зарайская И.Ю., Лукашев А.О., Редько В.Г. Проект «Мозг анимата»: разработка модели адаптивного поведения на основе теории функциональных систем // Восьмая национальная конференция по искусственному интеллекту с международным участием. Труды конференции. М.: Физматлит, 2002. Т.2. С.781-789.
58. Tsitolovsky L.E. A model of motivation with chaotic neuronal dynamics // Journ. of Biological Systems. 1997. V. 5. N.2. PP. 301-323.
59. 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. См. также http://www.keldysh.ru/pages/BioCyber/PrPrint/PrPrint.htm
60. Бурцев М.С., Гусарев Р.В., Редько В.Г. Исследование механизмов целенаправленного адаптивного управления // Изв. РАН "Теория и системы управления" 2002. N.6. С.55-62.
61. Бурцев М.С. Модель эволюционного возникновения целенаправленного адаптивного поведения. 2. Исследование развития иерархии целей // Препринт ИПМ РАН, 2002, N. 69.
62. Турчин В.Ф. Феномен науки. Кибернетический подход к эволюции. М.: Наука, 1993. 295с. (1-е изд). М.: ЭТС, 2000. 368 с. (2-е изд). См. также http://www.refal.ru/turchin/phenomenon/
63. Red'ko V.G., Prokhorov D.V., Burtsev M.S. Theory of functional systems, adaptive critics and neural networks // International Joint Conference on Neural Networks, Proceedings. Budapest, 2004. PP. 1787-1792.
64. Edelman, G. M. *Neural Darwinism: The Theory of Neuronal Group Selection*, Oxford: Oxford University Press, 1989.
65. Lehmann, F. (Ed). Semantic Networks in Artificial Intelligence, Pergamon Press, Oxford, 1992.
66. Воронин Л.Г. Эволюция высшей нервной деятельности. М.: Наука. 1977. 128 с.
67. Red'ko V.G. Evolution of cognition: Towards the theory of origin of human logic // Foundations of Science. 2000, Vol.5. N. 3. PP. 323-338.

Удалено: Сайт MIT

68. Редько В.Г. Пролегомены к теории происхождения мышления // Статья на сайте Круглого стола IV Всероссийской научно-технической конференции "Нейроинформатика-2002": http://www.keldysh.ru/pages/BioCyber/RT/Redko/Redko2.htm
69. Редько В.Г. Адаптивный сайзер // Биофизика. 1990. Т.35. Вып.6. С.1007-1011.
70. Staddon J. E. R. On rate-sensitive habituation // Adaptive Behavior. 1993. Vol. 1. N. 4. PP. 421-436.
71. Guillot A., Meyer J.-A. From SAB90 to SAB94: Four years of animat research // Cliff et al. (Eds). From Animals to Animats 3. Proceedings of the Third International Conference on Simulation of Adaptive Behavior. The MIT Press.1994. See also: http://animatlab.lip6.fr/index.en.html
72. Ляпунов А.А. О некоторых общих вопросах кибернетики // Проблемы кибернетики. М.: Физматгиз, 1958. Вып. 1. С.5-22.
73. Grossberg S. Classical and instrumental learning by neural networks // Progress in Theoretical Biology. 1974. Vol.3. PP.51-141.
74. Barto A.G., Sutton R.S. Simulation of anticipatory responses in classical conditioning by neuron-like adaptive element // Behav. Brain Res. 1982. Vol.4. P.221-235.
75. Klopf A. H., Morgan J. S., Weaver S. E. A hierarchical network of control systems that learn: modeling nervous system function during classical and instrumental conditioning // Adaptive Behavior. 1993. Vol. 1. N. 3. PP. 263-319.
76. Balkenius C., Moren J. Computational models of classical conditioning: a comparative study // C. Langton and T. Shimohara (Eds.). Proceedings of Artificial Life V. MIT Press, Bradford Books, MA.: 1998. See also: http://www.lucs.lu.se//Abstracts/LUCS_Studies/LUCS62.html
77. Редько В. Г. Моделирование когнитивной эволюции – естественный путь к искусственному интеллекту // Новости искусственного интеллекта. 2001. N. 2-3. С. 52-56.