

Н. М. АМОСОВ

Искусственный интеллект сегодня	
	Системы и модели
Восприятие и память	
Действия с моделями	
Взаимодействие моделей в интеллекте	
	Функциональный акт
	Сознание и подсознание
Интеллект уровня человека	
Искусственный интеллект в человеческом обществе	
На пути к интеллекту выше человеческого	Проект алгоритмической модели интеллекта

АЛГОРИТМЫ РАЗУМА

ВВЕДЕНИЕ

1. Введение

2. Основные понятия

3. Алгоритмы

4. Заключение

914911

Киев
НАУКОВА ДУМКА
1979

ВОЛОГОДСКАЯ
областная библиотека
им. И. В. Бабушкина

Возможно ли создать искусственный интеллект? Будет ли он способен к полноценному мышлению и творчеству? Кем предстоит ему стать — помощником или соперником человеческого разума? Эти вопросы давно уже обсуждаются многими учеными.

Свою точку зрения по ним высказывает академик АН УССР Н. М. Амосов. Автор известен своими работами в области моделирования мышления и поведения. В книге излагаются его идеи в их дальнейшем развитии. Анализируются возможные пути построения искусственного интеллекта. Подытожен опыт отдела биокibernетики Института кибернетики АН УССР по моделированию интеллекта и личности.

Рассчитана на широкий круг специалистов в области кибернетики, психологов, а также на всех интересующихся вопросами современной науки.

Ответственный редактор А. М. КАСАТКИН

Редакция научно-популярной литературы

30501-351
А БЗ-5-5-79 1502000000
M221(04)-79

© Издательство «Наукова думка», 1979

Предисловие

Искусственный интеллект можно определить как свойство цифровой вычислительной машины или сети нейроподобных элементов реагировать на информацию, поступающую на ее входные устройства, почти так же, как реагирует в тех же информационных условиях некоторый задуманный или конкретный человек. Машина и моделируемый человек одинаково по своим результатам распознают образы и ситуации, решают логические и другие задачи, принимают одни и те же решения в конфликтных ситуациях, то есть, кратко говоря, демонстрируют одинаковые результаты мышления. Если машина имеет достаточный набор гибких выходных органов, то машина и человек совершают одинаковые движения, что в целом приводит к одинаковым результатам в поведении. В идеальном случае эмоциональная окраска результатов мышления и поведения у машины-модели и человека — объекта моделирования — также должна быть одинаковой.

Подобным же модельным образом можно определить и интеллект коллектива либо целого общества людей. Устранение в результате обучения модели ошибочных реакций позволяет поставить задачу о построении искусственного интеллекта, более совершенного, чем интеллект человека. Вопросы создания «сверхинтеллекта» рассматриваются в данной книге.

Требование равных для машины и моделируемого человека результатов мышления и поведения является общепринятым среди ученых при определении искусственного интеллекта. Но относительно способов достижения этого равенства существуют разногласия. Некоторые ученые допускают, чтобы машина получала

те же результаты только по своим, более свойственным ей алгоритмам. К таким машинным алгоритмам относятся программы, основанные на последовательном опробовании многих вариантов решения по целесообразно выбранным критериям — принцип самоорганизации модели. Другие же ученые (к ним принадлежит и автор этой книги академик АН УССР Н. М. Амосов) считают, что алгоритмы, по которым получаются одинаковые результаты мышления и поведения у машины и моделируемого объекта, должны совпадать. Способ открытия таких алгоритмов интеллекта — эвристический. Он основан на разновидности имитационных методов моделирования — эвристическом моделировании, разработанном Н. М. Амосовым для моделирования разнообразных сложных систем.

Характерной чертой имитационных методов моделирования является то, что они исходят из общих представлений автора модели об объекте. Экспериментальные данные или опытная проверка не требуются принципиально. Так, например, при построении модели трудовой активности личности достаточно указать общий характер нелинейных зависимостей типа «труд—плата», «плата—чувство», «чувство—труд» и др.

Подобно тому как это делается при имитационном динамическом моделировании Дж. Форрестера, характеристики элементов системы задаются автором модели машине, которая, решая полученную ею систему уравнений при различных начальных и граничных условиях, демонстрирует на дисплее возможные «сценарии» — варианты происходящих в системе процессов. При этом подтверждается, что система вовсе не является простой суммой ее элементов, так же как и решение системы уравнений не является суммой решений каждого из них.

Эвристическое моделирование Н. М. Амосова, в отличие от динамического моделирования Дж. Форрестера, оперирует в основном графическими нелинейными характеристиками элементов, а не их дифференциальными уравнениями, хотя использование последних не исключается. Это допустимо, так как изучаются только постоянно действующие установившиеся зависимости (об ограничениях области рассмотрения, принятой в книге, будет сказано ниже). Наблюдающийся сейчас несомненный успех применения и быстрое распространение имитационных методов моделирования объясняются удачным распределением «обязанностей» между человеком и машиной. Человек хорошо придумывает поэлементные характеристики, а машина хорошо обобщает их в единую систему уравнений, решая которую получает искомые результаты. Слабой стороной имитационных методов моделирования является их субъективный характер. Если изменятся представления автора модели о характеристиках элементов объекта, то изменятся и модели. К тому же практически сколько авторов — столько и мо-

делей. Двух авторов с совершенно одинаковыми представлениями об объекте найти трудно. Здесь уместно напомнить читателю о существовании других, объективных методов моделирования, таких, как метод эвристической самоорганизации моделей, основанных на обработке небольшой таблицы опытных данных. Различие состоит в том, что имитационные методы моделирования, не требуя опытных данных, позволяют получить модели некоторых обобщенных объектов, тогда как экспериментальные методы дают модель конкретной личности, коллектива и общества.

В данной книге рассматривается имитационное моделирование указанных объектов в постановочном плане. На первых порах искусственный интеллект может не достигать полного совпадения модели и объекта. Моделирование вначале может решать только часть интересующих нас актуальных задач.

Основные ограничения следующие: динамика процессов выбора цели (целеполагания) и образования «установки» не рассматривается; цель объекта и система правил поведения считаются заданными или уже выработанными; эмоции не моделируются; рассматривается только неслучайная, регулярная составляющая психологических процессов; динамика не учитывается, поскольку уравнения элементов системы задаются как установившиеся характеристики.

В книге оговаривается, что эмоции человека в описываемых алгоритмах пока не моделируются. Кроме того, моделирование учитывает только так называемую регулярную составляющую результатов мышления и поведения, обусловленную воздействиями — «стимулами». При этом учитываются как точка зрения бихевиористических теорий (закон «стимул—реакция»), элементы гештальт-психологии (учет «врожденной», то есть заложенной при создании машины образной информации), так и предыстория одновременного взаимодействия многих элементов (обучение). Случайная составляющая результатов моделированию не подлежит. Поясним это. Мышление, как известно, связано с процедурой постановки задач и принятия решений. Каждое из решений происходит в условиях воздействия на личность многих признаков (термины «признак», «стимул» и даже «критерий» в данном контексте являются синонимами). В области задачи, где все признаки действуют согласованно, в одном направлении, решение является регулярным, однозначным, определенным. Например, если все рецепторы свидетельствуют о том, что распознаваемая буква есть буква А, то другая классификация буквы исключается. Отметим, что нас будет интересовать область случайных или нерегулярных решений задач, где действуют противоречивые признаки или стимулы. Грубо говоря, если сумма противоречивых стимулов с учетом их важности близка к нулю, например часть

7 рецепторов утверждает: «буква А», а другая часть — «буква В».

то мозг принимает «чисто» случайные, равновероятные решения. Именно в этой области проявляется так называемая свобода выбора. В книге указывается, что в случае нерегулярности, согласно учению о доминанте, происходит всемерное усиление внимания и разрешающей способности, но если и оно не помогает, то в действие вступает случай.

Предсказать случайное решение принципиально невозможно, так же как нельзя регулярно угадывать результат тиража «спортлото» или предвидеть, на какую сторону упадет подброшенная вверх монета. Область случайных решений поддается моделированию только в том смысле, что и в модели можно предусмотреть генератор случайных движений. Однако направление случайного движения регулярно угадывать нельзя. Все сказанное хорошо согласуется с известной теорией многокритериального управления Парето (1909 г.), в соответствии с которой в области эффективных решений, где критерии противоречат друг другу, математика, а следовательно, и моделирование бессильны. Здесь для принятия решений обычно привлекаются эксперты, которые знают, что же нужно выбрать. Машина может только существенно помочь выбору, используя дополнительно процедуру прогнозирования будущего для каждого из входящих в область решений Парето (журн. «Автоматика», 1978, № 2). Добавление процедуры прогнозирования можно использовать и для сокращения области случайных решений человека. Область, где моделирование возможно, при этом расширяется. Вопрос о принципиальной возможности моделирования сводится к выяснению соотношения области регулярности и области случайного выбора в актуальных задачах. Профессор В. В. Налимов, например, утверждает, что в интересующих нас задачах биологии и социологии участвует такое множество противоречивых воздействий, что моделирование невозможно (журн. «Автоматика», 1977, № 4). Конечно, это является преувеличением. Н. М. Амосов исходит из того, что область регулярных неслучайных решений при моделировании интеллекта достаточно содержательна, и ограничивает свое рассмотрение этой областью.

Привлечение идей Парето позволит в какой-то мере согласовать позиции сторон в научной дискуссии об экспериментально-вербальном, психологическом и математическом информационно-кибернетическом подходах в исследовании интеллекта. В частности, утверждение психологов о том, что в психике человека имеется принципиально не поддающаяся моделированию область решения задач (см., например, *Тихомиров О. К.* Развитие вычислительной техники и психологическая наука. — Вестник Московского ун-та. Психология, 1977, № 2, 3, 4), с учетом теории Парето получает ясное математическое объяснение. Экспериментальный характер моделирования обеспечивается рядом методов

кибернетического подхода, что также соответствует требованиям психологов.

Простейшим применением излагаемой в данной книге теории моделирования интеллекта является создание роботов, управляемых по вычислительным программам, содержащим проявления интеллекта. Созданный под руководством автора книги в отделе биокибernetики Института кибернетики АН УССР робот — трехколесная тележка ТАИР — обходит препятствия и находит оптимальный путь к цели. Наблюдения за интеллектуальными роботами пока являются своеобразным и, пожалуй, единственным методом экспериментальной проверки всей теории в целом: поведение роботов должно быть «интересным» и «умным».

Излагаемая в книге теория интеллекта в целом далеко выходит за пределы задачи управления роботами и предназначена для моделирования поведения и интеллектуальной деятельности как отдельной личности, так и коллектива или даже общественной системы. Робот в виде нейроноподобной аналоговой М-сети с положительной обратной связью только иллюстрирует возможность реализации значительно более общей идеи.

Н. М. Амосов смело ставит вопросы об имитационном объяснении на основе его моделей самых сложных вопросов психологии человека — таких, какие возникают при решении задач, исследовании подсознания, сновидений, «озарений» и других сложных явлений.

В заключение автор высказывает сомнение в том, удалась ли ему книга. Я уверен в ее большой пользе, в том, что она очарует не только кибернетиков, но и психологов, воспринявших имитационный метод моделирования, а также всех тех, кого интересуют тайны регулярного поведения живого и искусственного интеллектов. Что касается упомянутых выше значительных ограничений области исследования, то они будут поняты как необходимые при первопроходческом характере работы.

Член-корреспондент АН УССР
А. Г. ИВАХНЕНКО

Введение

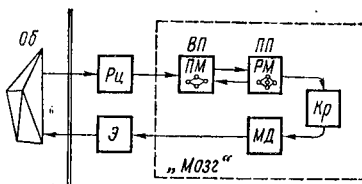
Механизмы разума интересуют ученых разных специальностей. Для психологов и физиологов — это теория их науки, для кибернетиков — пути создания искусственного интеллекта (ИИ). В предлагаемой вниманию читателя книге я попытаюсь дать изложение представлений об этой проблеме — как результат развития работ, ведущихся в отделе биокибернетики Института кибернетики АН УССР начиная с 1962 г. Новые идеи значительно отличаются от ранее опубликованных [1, 3].

Слово «алгоритм» не случайно введено в название книги: мне представляется, что есть возможность «разложить по полочкам» самые сложные проявления интеллекта — и даже с перспективой на его развитие выше уровня человеческого разума. Видимо, я не смогу убедить скептиков — для этого нужно воспроизвести алгоритмы интеллекта в программах. К сожалению, на этом пути стоят большие трудности. Может быть, излагаемые идеи как-то помогут энтузиастам проблемы? Предупреждаю, что предмет исключительно сложен для понимания, поскольку лежит на стыке физиологии, психологии, техники и даже философии. Для упрощения я буду широко пользоваться схемами.

Интеллект определяется как совокупность средств и способов управления сложными системами путем оперирования с их моделями, направляемого критериями

Рис. 1. Схема сетевого интеллекта (СИ):

Об — объект управления; Рц — рецепторы; Э — эффекторы; ВП — кратковременная память; ПМ — первичная модель; ПП — постоянная память; РМ — распознанная модель; Кр — критерии; МД — модель действия.



оптимальности управления. Современная наука и техника дают возможность воспроизводить модели и действия с ними техническими средствами и таким образом отделить разум от мозга, с которым его обычно связывают. Отличие приведенного определения от множества других состоит в том, что оно подчеркивает это последнее обстоятельство.

Таким образом, говоря об интеллекте, мы будем иметь в виду эту совокупность средств и способов управления, независимо от того, реализована ли она в биологических системах или при помощи искусственно созданных, технических средств. Такое употребление термина «интеллект» не является общепринятым. Оно, однако, тесно связано с основной идеей этой книги. В тех отдельных случаях, когда речь пойдет только о естественном интеллекте человека, это будет специально оговариваться. Что же касается термина «искусственный интеллект» (ИИ), то он, как это и принято, будет использоваться для обозначения различного рода технических реализаций, моделей интеллекта.

Остановимся на основных понятиях, с тем чтобы постепенно войти в круг обсуждаемых вопросов. Прежде всего нужно выделить два противоположных подхода к моделированию интеллекта. Условно их можно назвать сетевым и алгоритмическим. Соответственно будем различать и два типа моделей — сетевой (СИ) и алгоритмический (АИ) интеллекты.

Самая простая схема сетевого интеллекта показана на рис. 1. Объект управления (Об) представляет собой некую сложную трехмерную структуру. Она воспринимается датчиком — рецепторами Рц (например, глазами) и передается сигналами в «мозг», где превращается в первичную модель ПМ. В первом приближении ПМ представлена плоской двухмерной структурой, составленной из возбужденных, активированных элементов некоторой сети — условного «рецепторного поля». Объект распознается путем наложения ПМ на множество фигур из неактивных элементов, объеди-

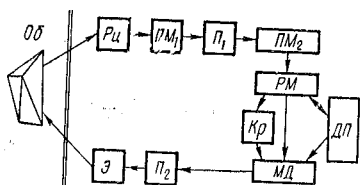


Рис. 2. Схема алгоритмического интеллекта (АИ):

Об — объект управления; Рц — рецепторы; ПМ₁ — первичная структурная модель; П₁ — преобразователь; ПМ₂ — первичная цифровая модель; РМ — распознанная модель; Кр — критерии; ДП — длительная память; МД — цифровая модель действия; П₂ — преобразователь; Э — эффекторы.

ненных «проторенными» связями. Эти неактивные модели-фигуры представляют собой модели уже известных объектов и составляют постоянную память ПП. В соответствии с принципом действия СИ элементы, составляющие постоянную и временную (или активную) память ВП, — одни и те же; они отличаются только уровнем активности. Наложением первичной модели на сеть выбирается и активируется одна распознанная модель — фигуры РМ, и от нее включается управляющее воздействие на объект. Это воздействие представлено моделью действия МД, управляющей эффекторами Э. Распознанная модель объекта связана с несколькими моделями действий; выбор одной, нужной, определяется критерием Кр. Активированная от РМ и Кр модель действий передает активность на эффекторы, в которых управляющие сигналы превращаются в механическую энергию управляющего воздействия. Таким образом, в сетевом интеллекте «действия с моделями» представлены изменением активности элементов сетей, в которых заложены модели.

На рис. 2 показана схема алгоритмического интеллекта. Начало его функционирования такое же, как в СИ, — в результате работы рецепторов Рц формируется первичная структурная модель ПМ₁ в виде такой же плоской двумерной структуры. Однако она тут же считывается преобразователем П₁, превращаясь в линейную, одномерную модель из набора цифр — ПМ₂. Все последующие действия осуществляются с этой моделью. Распознавание ПМ₂ состоит в последовательном сравнении ее с записанными тем же кодом моделями-эталоны из длительной памяти ДП, где находится распознанная модель РМ. По этой модели выбирается модель действия МД. Процесс этот осуществляется путем перебора моделей в длительной памяти под управлением критерия Кр. Модель действия передается на преобразователь П₂, где цифровой код

превращается в управляющие сигналы эффекторов Э и через управляющие действия — на объект Об.

Основное различие между СИ и АИ состоит в структуре памяти и вытекающих отсюда разных действиях с моделями. Однако в обоих типах интеллекта сохраняется принцип управления моделями со стороны критериев управления через их активацию.

Создание СИ и АИ предполагает использование разных методологических подходов к моделированию одного и того же объекта — разума человека. Каждый из подходов имеет свои сильные и слабые стороны. Поэтому при построении реальных систем ИИ может оказаться целесообразным представление некоторых функций разума в виде сетевых моделей, а некоторых — в виде алгоритмических. Общий принцип здесь таков: чем ниже уровень моделируемой функции в общей иерархии функций разума, тем вероятнее, что при ее воспроизведении в ИИ наиболее эффективным окажется сетевой метод.

Дело в том, что реализация многих функций нижнего уровня (как, например, нахождение распознанной модели по первичной модели — см. рис. 2) связана с осуществлением больших переборов в длительной памяти. Такие переборы особенно эффективно осуществляются в сетевых моделях, реализующих параллельные процессы переработки информации. В то же время сетевые модели имеют другие недостатки, ограничивающие область их применения. На сравнительном анализе достоинств и недостатков сетевых и алгоритмических моделей я еще буду неоднократно останавливаться.

Вопросам создания и исследования свойств СИ посвящен ряд работ, выполненных под руководством автора. Получены конкретные результаты, краткий обзор которых содержится в первом разделе книги. Все дальнейшее изложение я посвящу описанию принципов построения АИ, не делая больше специальных оговорок, что при практической его реализации ряд функций может быть представлен при помощи сетевых структур.



Искусственный интеллект сегодня

Основные этапы и направления исследований

Не буду даже пытаться охватить всю проблему искусственного интеллекта. Книга задумана как изложение собственной гипотезы об общих механизмах или алгоритмах интеллекта, которым равно подчиняются разум животных, человека, коллективный разум общества и которые, как мне кажется, обязательны для любой его модели.

Искусственный интеллект стал особой областью знания. Существуют специальные комитеты, координирующие исследования по этой проблематике, издается несколько журналов, созываются международные конференции. С одной стороны, предмет этих исследований примыкает к теоретической кибернетике, с другой — к технике автоматов и роботов, есть и психофизиологический аспект проблемы. Одно скажу сразу: искусственный интеллект не создан. Векселя, которые выдавала кибернетика в начале своего развития, претендуя на решение почти всех интеллектуальных задач, остались неоплаченными. Более того — наметился явный пессимизм во взглядах на саму возможность воспроизвести разум человека. Сошлюсь на книги Дрейфуса. Тем не менее большинство ученых-кибернетиков смотрит в будущее с надеждой, хотя теории интеллекта пока нет и все предлагаемые модели представляют собой воспроизведение частных механизмов мышления. Для того чтобы дать представление о состоянии дела, приведу очень краткую сводку работ последнего времени.

С самого начала основные надежды кибернетики были связаны с моделированием работы мозга. Было ясно, что здесь можно идти двумя путями — моделировать нейронные сети и воспроизводить алгоритмы мышления.

Лучшие результаты по моделированию нейронных сетей были получены в работах школы У. Мак-Каллоха [7]. Выяснилось, однако, что сети из формальных нейронов не способны воспроизводить сложные функции мозга. Неэффективными оказались также попытки использовать такие сети для управления роботами [9]. Большой интерес вызвали работы Ф. Розенблата, который сформулировал ряд принципов нейродинамики мозга и использовал их для построения персептронов — устройств для распознавания образов. Однако строгий анализ, осуществленный М. Минским и С. Пейпертом [8], показал ограниченность существующих здесь возможностей. Таким образом, к началу 70-х гг. общий кризис нейронного подхода стал очевидным.

Одновременно проводились исследования по алгоритмическому моделированию мышления. Основные достижения в этой области связаны с именами А. Ньюэла и Г. Саймона [12]. Их работы по созданию GPS (общего решателя проблем) привели к формированию отдельного направления — эвристического программирования, влияние которого на исследования по ИИ прослеживается до сих пор. Однако в целом это направление исчерпало свои возможности уже к началу текущего десятилетия. При этом выяснилось, что оно не позволяет приблизиться к сколько-нибудь общей теории мышления, хотя и может обеспечить решение отдельных прикладных задач.

Кризис обоих направлений привлек внимание к работам, начатым еще в 50-е гг. и связанным с использованием формальных методов для решения сложных, «интеллектуальных» задач (доказательство теорем, игры и т. п.). К этому времени здесь был получен ряд интересных результатов. Среди них следует отметить разработку А. Сэмюэлем [11] программы для игры в шашки, которая до сих пор является одной из лучших игровых программ. Работы этого направления и составили основу нового раздела кибернетики, который занимается проблемой искусственного интеллекта.

В начале 70-х гг. на исследования по ИИ оказывали большое влияние результаты, полученные в области математической логики Дж. Робинсоном. Развитый им

метод резолювенций позволил строить мощные процедуры доказательства теорем. Это было использовано для построения нового класса программ, решающих сложные задачи. Наибольшую известность среди них получила разработанная в Стенфордском исследовательском институте программа STRIPS [14], решающая задачи планирования действий робота. Методы математической логики составили также основу теории поиска решений, главные положения которой обобщены в работах Н. Нильсона. Однако использование этих методов для широкого круга задач показало их низкую эффективность, обусловленную в основном большим объемом и громоздкостью соответствующих программ для ЦВМ. Так что к 1975 г. наметилось разочарование и в этом подходе.

Попытки преодолеть трудности, сопутствующие разработке больших программ, привели к появлению новых методов автоматизации программирования и созданию специальных проблемно-ориентированных языков. Большие достижения здесь связаны с работами К. Хьюитта по построению языка PLANNER [15, 16], который послужил основой для целого ряда дальнейших разработок в этом направлении.

Использование новых языков программирования обеспечило прогресс в сравнительно новой для ИИ области — имитации речевого поведения человека. Первые значительные успехи здесь были получены Т. Виноградом [5], разработавшим систему диалогового управления роботом с помощью естественного языка. Начиная с 1973—1975 гг. проблемы естественной речи привлекают все большее число исследователей. Разрабатываются проблемы понимания [13], представления знаний, грамматического анализа и др. Проблема понимания активно исследуется также и в другой области ИИ — распознавания зрительных образов. Широко известны работы М. Минского, развивающего теперь новую теоретическую концепцию восприятия, связанную с представлениями о «системах фреймов».

В целом в последние годы опять намечается тенденция к использованию в программах ИИ знаний о поведении и мышлении человека, хотя сама задача моделирования человека прямо не ставится.

Меня больше всего интересует как раз эта задача — моделирование человека. Естественно было бы обратиться к нейрофизиологии и психологии. Именно эти науки должны описать физиологические механизмы

психических явлений и дать рекомендации по их моделированию. Возможно, работающим здесь ученым кажется, что они знают, как мыслят животные и человек, но я не мог уяснить этого из их книг. Не нашел я и возможности переложить их словесные формулировки на цифровые модели. Впрочем, и другие кибернетики тоже не смогли перевести физиологию мозга в модели.

Разумеется, никто не будет отрицать, что именно нейрофизиологам принадлежат основополагающие открытия, благодаря которым стали возможными попытки строить модели. Перечислю несколько основных.

Первое — учение И. М. Сеченова о возбуждении и торможении. Разный уровень активности нейрона — от полного «молчания» до максимального возбуждения — это важнейший механизм взаимодействия моделей. Открытие И. П. Павловым условных рефлексов и условных связей дало объяснение обучаемости и самоорганизации. Учение А. А. Ухтомского о доминанте позволяет предположить механизм превалирования главной модели над всеми остальными. Если к этому добавить работы о ретикулярной формации, то, вместе взятые, они подсказали идею центрального регулирования активности коры — систему усиления-торможения (СУТ) и механизм сознания и подсознания. З. Фрейд достаточно обосновал большое значение подсознания в психической жизни человека. Опыты Х. Дельгадо и Н. П. Бехтеревой со вживленными электродами показали, как центры чувств управляют поведением, являясь источниками активности для моделей коры. Наконец, исследования Н. П. Бехтеревой и др. обнаружили так называемый «код слов», то есть наличие самих моделей слов речи, которые, видимо, запечатлены в нескольких или даже многих ансамблях нейронов. Следует также упомянуть П. К. Анохина, предложившего общие принципы управления физиологическими процессами.

Простое перечисление основных идей нейрофизиологии показывает, что наша гипотеза об алгоритме интеллекта является лишь одним из возможных вариантов их компоновки в единое целое. Это сделано для того, чтобы попытаться применить метод эвристического моделирования для изучения мышления и поведения. Если нейрофизиология идет к психике «снизу» путем анализа, от механизмов нейронов и их ансамблей, то кибернетическое моделирование идет «сверху», путем синтеза. Мне представляется, что второй подход

914911

имеет право на научность. Не уверен, что модели уже сейчас предложат физиологии новые идеи для экспериментов, но для создания искусственного интеллекта они дадут достаточно материала. К сожалению, попытки заинтересовать моделями физиологов и психологов пока терпят неудачу. Они кажутся им слишком примитивными в сравнении с огромной сложностью мозга. Что ж, против этого пока нечего возразить. Нужны более доказательные модели, способные продемонстрировать как можно больше феноменов поведения человека. Именно к этой цели и направлены работы отдела биокибернетики Института кибернетики АН УССР. Приведу их краткую историю.

Некоторые результаты моделирования сетевого интеллекта

В 1963—1964 гг. мною был сформулирован первый вариант гипотезы о механизмах мышления в обобщенном виде. В 1965 г. эта гипотеза была опубликована в книге «Моделирование мышления и психики». Два года спустя книгу переиздали в США. Последующее развитие идей нашло отражение в монографии «Искусственный разум», напечатанной в 1969 г.

Модели, реализованные на ЦВМ. В 1963 г. мы начали серию работ по созданию моделей интеллекта, построенных на принципе семантических сетей с СУТ. Такие сети были названы термином «М-сети», а модели, построенные на них, — «М-автоматами». В последующие годы создавались новые варианты моделей, однако все они основывались на использовании сетевого интеллекта с СУТ. Принцип его действия уже был описан: М-сеть представляет собой систему элементарных моделей (i-модели), каждой придано определенное значение, например модель — объект среды, модель действия, чувства, отношения и др. Модели имеют статические и динамические характеристики, в которых отражена величина уровня активности и зависимости от времени и «входных» раздражителей — то есть от величины некоей энергии, поступающей по связям с других моделей или от рецепторов. У первых моделей связи задавались жестко, у последующих проходимость связей менялась в зависимости

от использования — тренировки. Это же относится к характеристикам возбуждения *i*-моделей. Исследование модели состоит в пересчете на ЦВМ уровня активности всех элементарных моделей за условный такт времени. СУТ выбирает одну, самую активную модель и еще дополнительно ее усиливает согласно своим характеристикам, а все остальные тормозит. «Входами» М-автомата являются внешние объекты, «выходами» — его действия, направленные на среду. Все это типично для СИ.

Первая модель — РЭМ отображала условный сюжет путешествия некоего «искусственного субъекта» в среде, которая содержала полезные и опасные для него объекты. Мотивы поведения субъекта определялись ощущениями усталости, голода и стремлением к самосохранению. Субъект изучал среду, выбирал цель движения, строил план достижения этой цели и затем реализовал его, выполняя действия-шаги, сравнивая результаты, получаемые в ходе движения, с запланированными, дополняя и корректируя план в зависимости от складывающихся ситуаций.

М-сеть РЭМа содержала 90 *i*-моделей, около 400 связей и была способна воспринимать три типа объектов (входных параметров). Среда могла содержать до 400 объектов, произвольно расположенных на плоскости. РЭМ выполнял восемь действий и был построен как неполный М-автомат, иными словами, кроме М-сети с заданной на ней СУТ, он содержал функционирующие сопряженно с М-сетью алгоритмические структуры. Большинство из них осуществляло функции планирования. РЭМ был реализован в виде комплекса программ для ЦВМ М-220. Программы содержали свыше 3000 рабочих команд. Время просчета одного такта функционирования М-сети (единицы автоматного времени) — 1,5 мин. В течение каждого такта производился расчет величин возбуждений и параметров характеристик всех *i*-моделей, а также параметров всех связей М-сети (эта процедура названа пересчетом). Последовательность операций, производимых за один такт, такова: восприятие информации из внешней среды, пересчет, выбор СУТ наиболее возбужденной *i*-модели, принятие на основании полученной информации решения о взаимодействии со средой, выполнение этого решения. Эксперименты с РЭМом включали просчеты его состояний и регистрацию решений на протяжении 10—20 тактов.

Результаты исследования РЭМа показали перспективность использования М-сетей как при построении моделей процессов мышления, так и при разработке систем типа «искусственный разум», способных к самостоятельному эффективному функционированию в сложных условиях. Выяснилась также целесообразность дальнейшего проведения работ по изучению возможностей аппарата при построении более мощных моделей, а также исследования весьма важного в практическом и теоретическом отношении вопроса об адаптивных возможностях М-автоматов.

Затем была построена и исследована новая модель — МОД. При его создании сохранялись как общая схема постановки задачи, так и принципиальная структура выбранного ранее условного сюжета моделирования. МОД также был разработан в виде неполного М-автомата. При этом его алгоритмическая часть отображала процессы планирования, связанные с предварительной организацией движения, а структурная — процессы принятия решений в ходе непосредственного взаимодействия со средой. Обе части целесообразно рассматривать как независимые модели мыслительных процессов, между которыми может быть организовано постоянное взаимодействие. Эти модели названы соответственно МОД-1 и МОД-2.

МОД-1 вырабатывал план деятельности М-автомата в среде в виде совокупности подцелей движения, основной цели и ожидаемых (предвидимых) состояний автомата в ходе движения. Алгоритмы восприятия внешней информации в МОД-1 моделировали процессы зрительного восприятия человека, построения возможных вариантов плана, их оценки, синтеза окончательного варианта, его разбиения на отдельные этапы, выбора объектов-ориентиров для каждого из этапов, моделировали процессы принятия решения человеком в условиях отсутствия полной информации о среде.

МОД-1 являлся неполным М-автоматом, основной объем его функций реализовался алгоритмическими структурами. М-сеть здесь содержала 62 i-модели, около 1000 связей и использовалась в основном для представления мотивационной сферы модели. МОД-1 был реализован в виде комплекса программ для ЦВМ М-220. Программы включали до 15 000 команд, их просчет занимал 20—30 мин машинного времени. Среда модели могла содержать до 625 объектов (это

максимальное число). Каждый из объектов задавался упорядоченным набором из 8—10 признаков (входных воздействий). Выходом модели являлся оптимальный план передвижений в среде. В среднем такой план мог состоять из 30 элементов (целей, подцелей и ожидаемых отклонений внутренних состояний).

М-автомат МОД-2 вырабатывал конкретные реализации планов, построенных МОД-1. Его основная особенность связана с реализацией на М-сети программ самообучения. Последнее осуществлялось путем изменения веса первоначально заданных связей, установления новых связей и порождения новых узлов М-сети. Протекание процессов изменения структуры сети определялось особенностями среды, в которой действовал автомат, соотнесенными с поставленными перед ним задачами. Задачи могли формулироваться, например, в следующем виде: действовать так, чтобы обеспечить максимальное значение оценки «собственного комфорта», максимальное соответствие внешних и внутренних реакций автомата реакциям моделируемого объекта или быстрее достижение цели. Возможны были и комбинированные задачи.

Автомат может рассматриваться как модель деятельности человека по принятию решений в задачах движения в лабиринте. Кроме того, МОД-2 может быть использован в качестве устройства, управляющего передвижением технических систем, предназначенных для сбора информации, транспортировки и т. п. В зависимости от цели использования автомата изменяется и критерий оценки эффективности его функционирования.

МОД-2 — полный М-автомат. Его М-сеть может содержать до 400 i-моделей и до 2000 связей. На М-сети задана двухуровневая СУТ. МОД-2 был реализован в виде программы для ЦВМ БЭСМ-6. Программа содержала около 6000 команд. Время просчета одного такта 30—50 сек. Экспериментально были исследованы реакции автомата на протяжении до 300 тактов. На входы автомата подавалась информация о плане движения и об объектах среды, каждый из которых относится к одному из шести возможных типов. Всего среда могла содержать до 400 расположенных произвольным образом объектов. Предварительная информация о некоторых характеристиках среды и структуре оценочных функций задавалась в процессе начальной организации М-сети автомата. Количество

выполняемых МОД-2 действий (выходных параметров или решений) — 22. Из них 17 — различного рода действия — шаги, перемещающие автомат в среде, а 5 — «активные» действия, изменяющие состояние среды или автомата («есть», «спать», «нести объект», «бросить объект», «создать убежище»). Поведение автомата состояло в формировании последовательностей решений о выполнении тех или иных действий и соответственно их фактическим выполнением. Были разработаны оценки поведения и предложены процедуры оптимизации автомата. Оптимизация выполнялась варьированием значений шести параметров обучения.

В ходе исследования МОД-2 был решен ряд вспомогательных задач-тестов. Представляет самостоятельный интерес результат одной из таких задач, связанной с моделированием процессов формирования понятий человеком. Здесь показана приводимость М-автомата к формам, моделирующим как индивидуальное, так и обобщенное групповое поведение. Были продемонстрированы методы такого приведения.

Модель механизмов речи. Одновременно с разработкой МОДа проводилось исследование, цель которого состояла в том, чтобы изучить возможности М-сетей в области нейрофизиологии и нейропсихологии, а также оценить практическую и познавательную важность таких моделей. Был разработан и исследован М-автомат, моделирующий механизмы речи. В модели представлены такие аспекты устной речи, как восприятие, осмысливание, словесное выражение. Преимущественное внимание уделялось содержательной стороне процессов переработки словесной информации. Модель предназначена для воспроизведения относительно простых речевых функций — ответов на вопросы ограниченного типа, повторения, называния. Она содержит следующие блоки: слуховых восприятий, сенсорный речевой, проприоцептивный речевой, понятийный, эмоций, мотивационный, двигательный речевой, артикуляторный и блок СУТ. Блоки модели соотнесены с определенными мозговыми образованиями. При задании организации М-сети использовались данные нейроморфологии, нейрофизиологии и клинической неврологии.

Модель представлена в виде необучающегося полного М-автомата. Его М-сеть содержит более 1000 i-моделей и 8000 связей между ними. М-автомат реализован в виде программы для ЦВМ БЭСМ-6, содержащей около 500 команд. Время просчета одного так-

та — 2 сек. В экспериментах наблюдалось поведение модели на протяжении до 100 тактов. На вход модели подавались буквы русского алфавита, объединенные в слова и фразы, а также специальные объекты, соответствующие образам предметов. На выходе модели, в зависимости от режима ее работы, наблюдались последовательности букв русского алфавита, которые были либо ответами на входные вопросы, либо повторением входных слов, либо названиями предметов. То обстоятельство, что при создании модели широко использовались данные нейрофизиологии, позволило в экспериментах имитировать ряд поражений мозга органического и функционального характера, приводящих к нарушениям функций речи. В частности, получены модельные отображения синдромов сенсорной, моторной, проводниковой и транскортикальной афазий.

Описанные М-автоматы составляют основной фонд «больших» моделей, разработанных и исследованных в процессе изучения возможностей и практических методов использования М-сетей.

Нами были выполнены и некоторые модельные разработки, в которых аппарат М-сетей использовался эпизодически или в модифицированном виде. Анализ результатов, полученных в разработках такого рода, может представлять серьезный интерес при оценке возможностей и свойств обсуждаемого нами аппарата.

Весь опыт моделирования поведения «разумного субъекта» в некоей среде — «лабиринте» — с использованием М-сети и расчетами на ЦВМ подытожен в монографии [4]. В качестве примера на рис. 3 приведены результаты одного из экспериментов по исследованию поведения такого «субъекта».

Семь-восемь лет мы занимались созданием сетевых моделей на ЦВМ, пока не убедились, что возможности таких моделей ограничены. Объем расчетов оказался слишком большим даже для компьютера: за один такт нужно пересчитать циркуляцию «энергии» по всем связям и изменения в их проходимости, подсчитать активность всех элементарных моделей, пересчитать изменение их тренированности для следующего такта. Если же предусмотреть и возможность образования новых связей и новых моделей, иными словами, воспроизвести принцип самоорганизации, то количество счетной работы будет расти подобно снежному кому. Затраты машинного времени увеличиваются прибли-

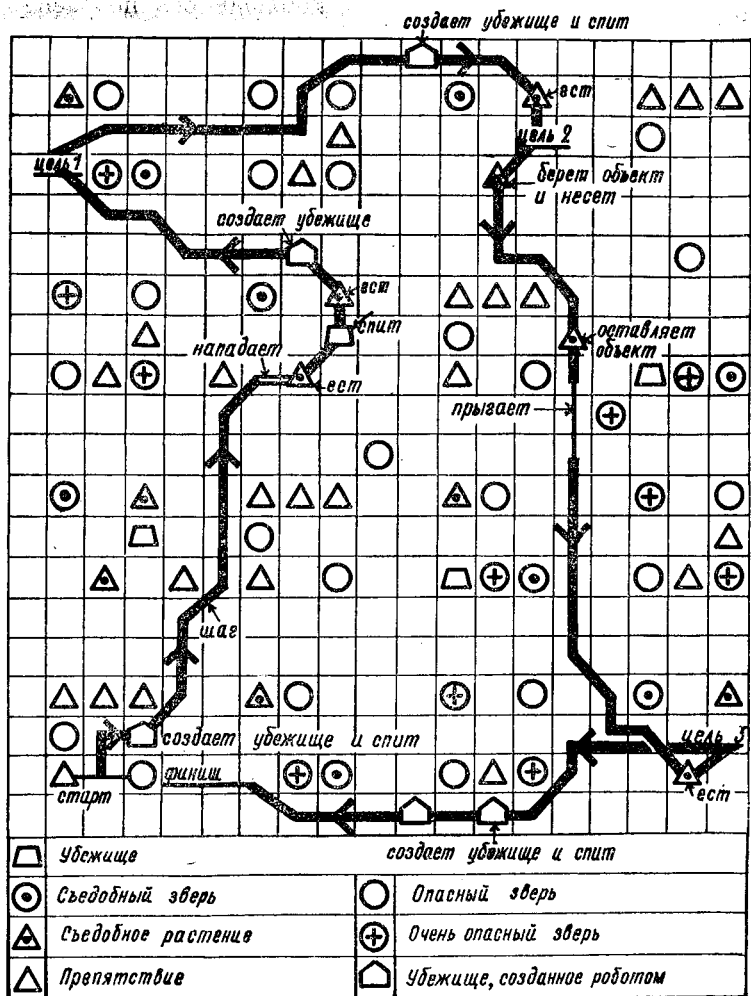


Рис. 3. Схема движений и действий МОДа по карте.

зительно пропорционально кубу числа моделей в сети. Но дело не только в расчетах — так же трудно оказалось отладить громоздкие программы. Так или иначе, выйти за предел 1000 моделей и 8000 связей нам не удалось. Формально мы воспроизвели в МОД самые простые программы психики, такие, как сознание и подсознание, оптимизация действий по многим крите-

риям — чувствам с предвидением и планированием. Была продемонстрирована разная обобщенность или иерархия моделей, обучение, забывание и даже различия характера. Но в целом этот «субъект», путешествующий среди врагов и препятствий в поисках пищи, соответствовал лишь довольно примитивному животному.

Аналоговые модели. Роботы. Тем не менее существует много задач, для которых вполне достаточен и такой ограниченный интеллект. В частности — для роботов, предназначенных для специализированной деятельности. Важнейшим требованием для них должна быть автономность, независимость от ЦВМ, что привело к реализации сетевого интеллекта на физических элементах. Идея сама по себе проста: представить каждую элементарную модель в виде усилителя, на вход которого поступает потенциал от других моделей, а на выходе формируется усиленный потенциал, который тоже передается по связям и гасится пропорционально их сопротивлению. Из таких элементов-усилителей можно создать любую сеть, если каждому придать определенное значение — семантику. Одни элементы — модели предметов, другие — чувств и т. д., как в сетевых моделях РЭМ и МОД. Разные характеристики усилителей и разные сопротивления связей позволяют создать структуры любого назначения. Модель интеллекта на физических элементах в наибольшей мере приближается к имитации мозга. К сожалению, есть разница: несоизмеримо мало число элементов и связей. Однако сложность такого интеллекта целиком определяется технологией. Можно создать довольно большие сети, во всяком случае достаточные для робота.

В 1972 г. были начаты работы по созданию модели ИИ в виде сети из физических элементов. Цель их — построить систему управления в виде М-сети с системой усиления-торможения (СУТ) и разместить ее на тележке.

Прежде чем приступить к разработке макета, были проверены возможности сети из физических элементов с СУТ. Первая сеть содержала 26 узлов и около 300 связей и управляла перемещением писчика по условной среде, изображенной на карте. Результаты экспериментальных исследований этой системы оказались обнадеживающими. Тогда приступили к созданию

Конструктивно разрабатываемый макет робота представляет собой трехколесное шасси, на котором смонтированы комплект рецепторов (органов чувств), блок управления, энергосистема и прочие устройства. Размер шасси $1600 \times 1100 \times 600$ мм. Все три колеса являются ведущими и имеют автономный привод от электродвигателей мощностью по 30 Вт. Переднее колесо поворотное. Питание электродвигателей осуществляется от аккумуляторных батарей. Скорость движения по ровной поверхности составляет 10—12 м/мин.

Датчики-рецепторы можно разбить на несколько групп.

1. Датчики, определяющие положение робота в пространстве:

а) навигационная система с компасом и двумя радиомаяками;

б) датчики углов наклона тележки в двух плоскостях.

2. Датчики информации об окружающей среде:

а) дистантные. Активный оптический дальномер с радиусом действия до 10 м. Система оптических датчиков близости с диапазоном расстояний до 30 см;

б) контактные датчики — система микровыключателей, установленных на гибком чехле, в который заключена тележка.

3. Датчики состояний робота:

а) термодатчики на электродвигателях;

б) датчики крутящего момента на приводах к колесам;

в) датчики напряжения на аккумуляторных батареях;

г) вибродатчик.

4. Датчик времени.

Основу системы управления представляет физически реализованная М-сеть. Специфическая СУТ в М-сети задает положительную обратную связь по возбуждениям узлов и обеспечивает тем самым доминирование в каждый момент времени одного или нескольких узлов над всеми другими. Ввод и вывод информации в М-сети соответствуют возбуждению определенных ее узлов (входных и выходных).

В настоящее время система управления роботом предполагает осуществление целенаправленного движения с обеспечением собственной безопасности (объезд препятствий, избегание опасных мест, поддержание внутренних параметров в заданных пределах) и

минимизацию временных и энергетических затрат. Вся сеть, состоящая из 100 узлов, разбита на шесть сфер.

Сферы оценок и распознавания ситуаций являются входными. Аналогично сенсорным системам мозга человека здесь осуществляется анализ воспринятой датчиками информации, на основе которого выполняется интегральная оценка среды, условий задачи и собственных состояний.

Выбор характера поведения в текущей ситуации производится в сфере решений. При этом может быть принято решение, определяющее направление движения, осуществление какого-то сложного маневра или даже выполнение некоторого элементарного действия.

Организация самого двигательного поведения осуществляется тремя сферами, являющимися выходными для сети, — сферами маневров высшего уровня, маневров нижнего уровня и элементарных действий. Здесь формируется последовательность команд, поступающих к эффекторной системе — системе управления поворотным и тяговыми электродвигателями.

Все узлы сети (i-модели) представляют собой усилители постоянного тока со специальными характеристиками. Узлы каждой из выделенных сфер имеют свои определенные характеристики. Точно так же в каждой сфере действует своя СУТ. Связи между узлами выполнены из резистивных элементов в устройстве матричного типа. Для ввода информации от датчиков в сетевое устройство управления используется 60 каналов.

Внешний вид ТАИРа представлен на рис. 4. В настоящее время мы продолжаем работы по созданию новых систем управления роботами. Эти системы пока не претендуют на достижение высокого уровня интеллекта. Правда, для ограниченных целей нейроподобная семантическая сеть на физических элементах с СУТ хорошо себя зарекомендовала и привлекла внимание исследователей и конструкторов, которые думают об автономных роботах. Планы отдела биок cyberнетики по совершенствованию автоматов предполагают реализовать обучение, создать ансамблевую организацию из элементов, позволяющую резко увеличить количество моделей при том же числе элементов, улучшить зрительное восприятие среды и даже дойти до человеческой речи. Я не уверен, что эти планы воплотятся

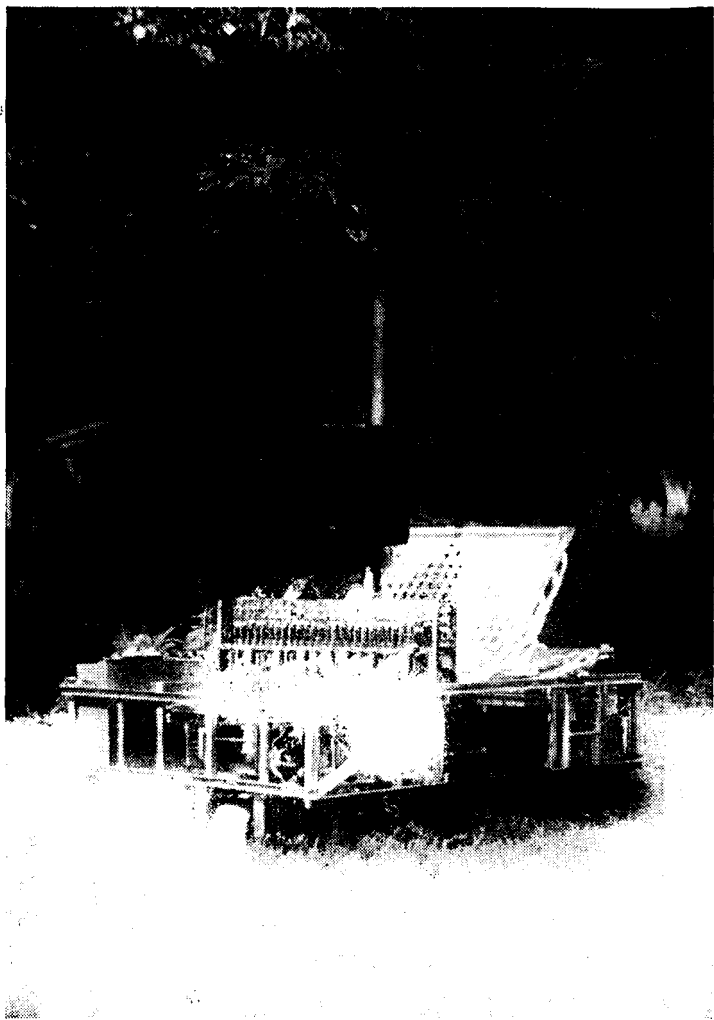


Рис. 4. Внешний вид ТАИРа.

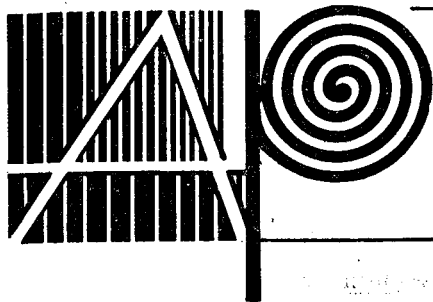
состоятельны, а исключительно вследствие трудностей их технической реализации. Почти пятнадцать лет попыток моделирования разума человека привели меня лишь к ответу на вопрос «Что такое интеллект?», но существенно не приблизили к созданию его модели. Нейроноподобная семантическая сеть с переменной 28

активностью элементов связей и с СУТ кажется наилучшим аналогом коры мозга, но размеры ее жестко ограничены сложностью воспроизведения, элементарные поведенческие реакции, полученные на описанных моделях, весьма далеки от человеческих и ничего не могут доказать. В самом деле, разве несколько сот элементов нашей сети могут заменить 10 миллиардов нейронов мозга? Если к этому прибавить, что каждая нервная клетка представляет собой функциональную систему из многих тысяч макромолекул, что она имеет несколько сот «входов» и может участвовать в работе различных клеточных ансамблей, давая почти астрономическое число моделей, то разве можно все это воспроизвести? К этому нужно добавить самоорганизацию в виде избирательной тренировки «входов» — синапсов, обеспечивающих память, и тренировку «выхода», резко повышающую активность нейрона. Нервная система — не просто сеть из одинаковых элементов. В ней имеется начальная структура связей и клеток с разной активностью, обеспечивающих врожденные рефлексы, чувства, программу доминирования. Тренируемость клеток и связей позволяет развивать и адаптировать врожденные реакции и наслаивать на них иерархию функциональных актов различных уровней обобщения и содержания. Все это, вместе взятое, дает человеческий разум — изумительное произведение природы, которое, раз возникнув, открыло в мозге новые возможности.

После такой характеристики сложности и возможностей мозга задача его воспроизведения в модели кажется безнадежной. Трудно представить себе искусственную сеть из десятка миллиардов элементарных усилителей, каждый из которых может иметь сотни входов, обладает способностью к тренировке — то есть изменению характеристик. Трудно, но не безнадежно. В отличие от длительной естественной эволюции прогресс науки и техники стремителен и все более ускоряется. Поэтому в перспективе возможна и аналоговая сеть, сравнимая по мощности с мозгом. Важно правильно поставить задачу — в данном случае сказать, какими должны быть элементы и как их соединять друг с другом. Пожалуй, еще важнее представить алгоритм интеллекта в достаточно обобщенном виде, позволяющем реализовать его различными средствами.

29 Велики технологические трудности на пути до аналогового интеллекта. Поэтому так заманчивы уни-

версальные цифровые машины, которые уже теперь достигли большой мощности. Совершенствуется их внешняя память и растет объем оперативной памяти. Быстродействие исчисляется миллионами операций в секунду. Разделение времени и создание параллельных программ позволяют повысить эффективность компьютеров. Создается впечатление, что возможности ЦВМ еще не достаточно использованы для реализации алгоритма интеллекта. То обстоятельство, что до сих пор наши попытки создания СИ не увенчались успехом, еще не означает, что исследования закончены. Нужно сохранить обобщенный алгоритм, но отказаться от сети с тем, чтобы уменьшить объем расчетов. Однако при этом следует лишь в минимальной мере поступиться принципами. Мы решили предпринять такую попытку — создать алгоритмическую модель интеллекта. Соображения к ее проекту будут представлены в заключительной части книги. А пока перейдем к изложению идей, положенных в основу этой модели.



Системы и модели

Современный этап развития науки ознаменован достижением принципиальной важности: вычислительные машины дали возможность овладеть сложностью. Все значение этого достижения как раз и состоит в том, что появилась надежда на создание количественных моделей, приближающихся по сложности к биологическим системам. Возможно, что при разработке таких моделей недостаточно внимания уделяется значению пространственных структур объектов, хотя наблюдения природы указывают на их исключительную роль (вспомним двойную спираль молекулы ДНК).

Весь физический мир можно свести к пространственным структурам, состоящим из атомов и молекул, а также к действующим между ними силам, связывающим материальные частицы в комплексы, которые условно можно назвать «телами». Общеизвестно, что все объекты в мире взаимосвязаны, однако степень прочности этих связей весьма различна: от жестких связей внутри твердых тел до гравитационных и электромагнитных сил, лишь ограничивающих пространственную свободу частиц. Пространство, энергия и время — вот самые общие координаты частиц и тел. Еще недавно казалось, что энергетические и материальные взаимоотношения между частицами и телами достаточно объясняют мир. Но вот появилось понятие информа-

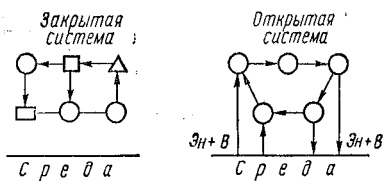


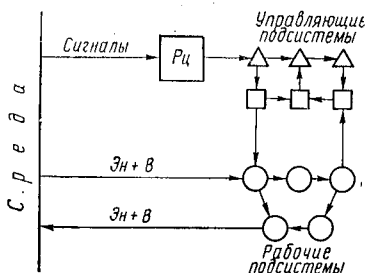
Рис. 5. Схемы простых систем — закрытой и открытой. Между элементами происходит обмен энергией $\mathcal{E}_n$ и веществом B .

ции и понятие сигнала как носителя информации, и это поколебало представления об исключительно энергетическом и материальном характере балансов отношений между объектами. Сигнал, несущий ничтожное количество энергии, может вызвать огромные вещественные и энергетические пертурбации в сложной системе, на которую он направлен (пример — атомная война).

Сложность структурных и энергетических отношений стала самостоятельным и значимым понятием, без учета которого уже невозможно объяснить мир. Понятие системы тоже более или менее определилось: это пространственная структура из неких элементов, объединенных внутренними «силами» настолько прочно, что она выступает как единое целое, противопоставленное всем другим объектам. Системы зависят друг от друга в обмене энергией и веществом, но в меньшей степени, чем элементы внутри них (рис. 5). Хотелось бы дать количественное понятие системы, но очень нелегко установить, когда простое сочетание взаимодействующих элементов уже становится системой. Степень «зрелости» системы условно можно определить по степени зависимости ее элементов друг от друга: сколько времени они могут «самостоятельно прожить», не распадаясь на более простые частицы, если их отделить от системы. В связи с этим понятие элемента системы тоже не просто — в конце концов все объекты разложимы до элементарных частиц. Мне кажется, что элементом системы нужно считать некое более простое образование, уже обладающее чертами данной системы. Если взять живые биологические объекты, то можно перечислить иерархические ступени их сложности: элементарные частицы, атомы, молекулы, макромолекулы (ДНК, белки), клетки, органы, организмы, сообщества, биоценозы... Каждый уровень сам по себе достаточно сложен по структуре, чтобы претендовать на звание «сложности», но все-таки,

Рис. 6. Схема сложной системы:

P_c — рецензор; $\mathcal{E}_n$ — энергия; B — вещество.



какие из них допустимо считать сложными, какие отнести к простым, какие признать лишь элементами сложных? Без условности здесь не обойтись.

Сложные системы

Будем считать сложными такие системы, в которых между элементами циркулируют не только частицы вещества и энергии, но и сигналы (рис. 6). В структуре сложных систем можно условно выделить рабочие подсистемы, ведающие преобразованиями вещества и энергии, и управляющие, которые воздействуют на рабочие с помощью сигналов. Хотя сигнал тоже имеет физическую, то есть вещественную и энергетическую, природу, но дело не в ней, а в характере сигнала, то есть его временной структуре и, особенно, месте приложения к управляемому объекту — в данном случае к рабочей подсистеме. При таких условиях — обязательность наличия управляющих сигналов и рабочих подсистем — грань сложных систем проходит на уровне одноклеточных существ: их управляющим органом является генетический аппарат ДНК, рабочими подсистемами — органеллы клетки (оболочка, митохондрии, лизосомы и др.). Роль сигналов выполняют информационные РНК. Макромолекулы — белки и нуклеиновые кислоты — достаточно сложны по структуре, но не удовлетворяют требованиям, предъявляемым к сложной системе. На более высоких уровнях иерархии систем эти условия соблюдены. Например, в организме органами управления являются нервная и эндокринная системы, сигналами — молекулы гормонов и медиаторов (передатчики нервных импульсов). Сообщество животных не

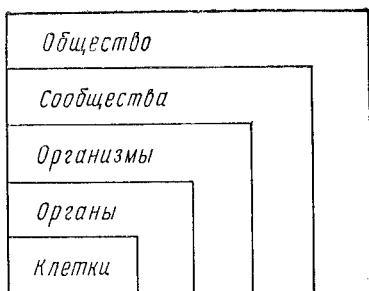


Рис. 7. Иерархия сложных систем.

всегда становится сложной системой. Только у высших млекопитающих и птиц есть внутренняя организация в стае и система управляющих сигналов, и только у человека эта система приобретает достаточную «зрелость». В обществе легко обнаружить структуры, аналогичные рабочим и управляющим подсистемам, в нем циркулируют многочисленные и разнообразные сигналы. Иерархия сложных систем представлена на рис. 7.

Элементом сложной системы каждого уровня являются системы предыдущего уровня, в которых уже заложены некоторые качества высшей системы. Для организма — это клетки, для общества — люди. Элементом клетки являются макромолекулы. Они способны воспроизводить себя лишь при наличии ферментов, действующих извне. Именно поэтому макромолекулу нельзя считать сложной системой. Скачок от молекулы до клетки очень велик — этим определяются трудности объяснения возникновения жизни на Земле. На других, высших уровнях такие качественные скачки менее выражены. Клетки многоклеточных, будучи отделены от тела, способны еще некоторое время жить, так же как и отбившиеся от стаи животные. А уровень «зрелости» такой системы, как человеческое общество, возрастает буквально на наших глазах. Еще пять — десять поколений тому назад, когда преобладало натуральное хозяйство, большинство людей было способно существовать в условиях весьма ограниченных связей с обществом. Теперь же брошенный в лесу человек может погибнуть через несколько дней.

Самое общее качество сложных систем «типа живых» — способность к поддержанию своей целостности. 34

ности и к противодействию разрушающим влияниям окружающей среды. Однако оно не беспредельно, поэтому необходимо другое качество, более сложное в своем структурном выражении,— способность к воспроизведению самих себя. Еще более сложным качеством является способность к усложнению в процессе воспроизведения. В живой природе это выражается изменчивостью. В человеческом обществе усложнение структуры и функции наблюдается постоянно и является следствием феномена творчества и труда, отсутствующих в стае животных.

Принцип структурности предполагает, что для реализации всех этих качеств должны быть соответствующие структуры. Нужна структура для постоянного возобновления своих разрушающихся частей и для утилизации с этой целью энергии среды, нужны структуры для размножения и структурное выражение программы их «удвоения» и, наконец, необходимы некоторые структурные возможности для наращивания новых структур, то есть для усложнения. Более того, должны быть структуры, отражающие внешний мир, поскольку на него замыкается реализация программ, которые являются выражением названных «способностей». Программа творчества тоже требует структурного выражения.

Модели

Существует хороший термин для обозначения структуры, отражающей другую структуру. Этот термин — модель; у каждого из нас есть его интуитивное понимание. Модель отражает объект не полностью, а с упрощениями и искажениями — в зависимости от того, для какой цели она предназначена и какие есть возможности для ее построения.

Органы управления сложных систем «типа живых» содержат в себе модели и программы (иными словами, тоже структуры), управляющие рабочими подсистемами в соответствии с этими моделями. Программа — это «считывание» модели сигналами, которые регулируют потоки энергии и вещества между элементами системы.

35 Как объяснить, что сравнительно простая структура — модель в ДНК зародышевой клетки человека — может отразить всю сложность самого человека с его

разнообразием клеток и всеми человеческими качествами? Для того чтобы из яйцеклетки вырос человек, нужно извне получить массу сложных веществ («кирпичиков»), а модель должна только предусмотреть, как сложить из них «здание». Само «складывание» состоит в значительной мере в повторении одинаковых операций. В ДНК заложены структуры всех белков организма и порядок, в каком следует «считывать» их при построении органов. Заложены и обратные связи, отмечающие выполнение этапов формирования организма. Следовательно, принцип управления по модели не исключает возможности построения более сложной, чем сама модель, системы, поскольку в модели должен быть предусмотрен лишь порядок включения структур, получаемых извне.

Мы привыкли к статическим моделям: игрушка, чертеж, текст — все это чистая структура. Основу моделей в органах управления сложных систем тоже составляют структуры: ДНК, сети нейронов в центральной нервной системе. Несомненна избыточность этих структур — например в каждой клетке имеется полный набор генов, достаточный для построения целого организма. Для реализации управления нужна программа «считывания» структур сигналами, а для этого — активация определенных частей модели. Таким образом, важна не только структура, но и активность, энергия различных элементов модели. Понятие активности тоже можно свести к изменению структур, только на уровне более низком, например в молекулах, атомах, составляющих структуру. Сам сигнал представляет собой такую «активированную порцию структуры». Пример — информационная РНК или активность синапса на теле нервной клетки, когда на него подается импульс с другого нейрона.

Структура нам представляется чем-то стабильным, хотя в действительности это относится лишь к грубым материальным конструкциям из многих молекул и атомов. В живых системах структуры ДНК очень прочны, чего нельзя сказать о структурах нервной системы. Конечно, нейроны не передвигаются, отростки их у взрослого растут медленно, но тонкие структуры синапсов — мест соединения нейронов, обеспечивающих прохождение энергии с одного нейрона на другой, — довольно нестабильны во времени. То же касается и «рабочих» элементов нейрона, обеспечивающих его активность — процесс возбуждения.

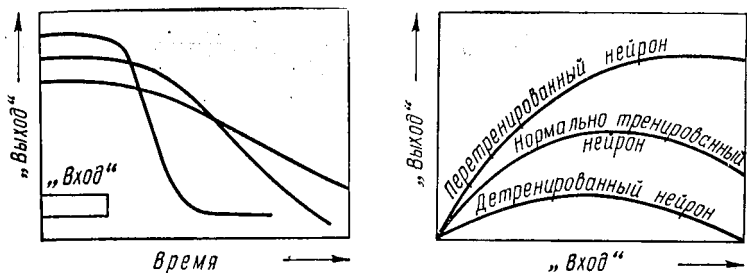


Рис. 8. Динамические характеристики элементов модели (нейронов):

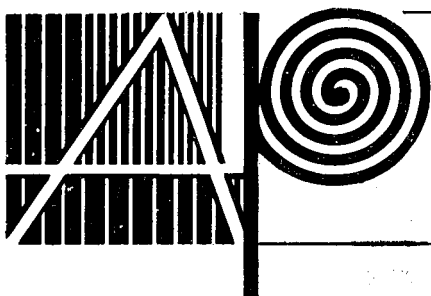
«вход» — время раздражения извне; «выход» — величина активности на выходе элемента. Характеристики отмечают длительность активности после прекращения действия «входа».

Рис. 9. Статические характеристики элементов модели (нейронов):

«вход» — величина раздражителя; «выход» — уровень активности. Показаны три характеристики — в зависимости от степени тренированности элемента.

Активность элементов модели, так же как и сигналы, — это ее функции. Модель, в которой элементы взаимодействуют друг с другом и с внешней средой через сигналы, можно назвать действующей, в противоположность статичной, лишенной функции и возможности самостоятельно взаимодействовать с внешним миром. Структуры живых клеток подчиняются закону тренировки и детренировки: при функционировании их «мощность» возрастает, при покое — уменьшается. Эти процессы развиваются неравномерно. Уровень и длительность активности живого структурного элемента не только заложены в его генах, но и являются результатом тренировки в процессе предшествовавшей деятельности. Изменение структуры модели в результате получения сигналов извне составляет память. Временная активность комплекса структурных элементов модели — это временная или активная память. Организация новых структур в соединениях элементов модели — ее длительная или пассивная память. На рис. 8 показаны временные (динамические) характеристики различных типов элементов, а на рис. 9 — статические характеристики их тренированности.

Итак, сочетание постоянных и изменчивых структур, состоящих из элементов с разными динамическими характеристиками активности, — вот черты моделей систем «типа живых».



Восприятие и память

Восприятие и управление

Управляющие органы сложных систем можно представить как их интеллект, «тело» — как рабочие подсистемы. Сигналы представляются комплексами подвижных структур, как, например, РНК в клетках или импульсы в нервных проводниках.

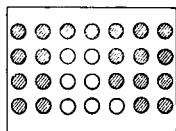
В связи с этим можно говорить о «коде» моделей и сигналов.

Под словом «код» понимается определенный набор структурных элементов, из которых строятся статические модели. Письменная речь — это код букв. Есть код рисунков, код математических формул. Структура клеток и организмов записана генетическим кодом — набором генов, составленных из нуклеотидов и объединенных в ДНК. Структурный код нервной системы — это нервные клетки различного типа.

Функция моделей выражается сигналами, и к ним тоже применяется понятие кода. Примеры — код нервных импульсов, код информационных РНК, код сигнальных флагов и масса других.

Управление любым объектом предусматривает цепь из трех звеньев: рецептор — датчик, воспринимающий энергию объекта и превращающий ее в сигналы, сами органы управления (моделирующая установка, мозг, разум, интеллект) и эффектор — орган воздействия, превращающий управляющие сигналы, исходящие от интеллекта, в некоторый вид энергии, достаточно мощ-

Рис. 10. Схема кратковременной памяти. Модель представлена кратковременной высокой активностью элементов (нейронов), показанных в виде заштрихованных кружочков.



ный, чтобы изменить деятельность объекта. Рецепторы специфичны: рецептор одного вида воспринимает только один вид энергии; но разные рецепторы одной системы превращают их в знаки одного кода, характерного для данного управляющего органа, интеллекта, разума. Например, рецепторы глаза, уха, кожи — все кодируют свои сигналы нервными импульсами. То же касается и эффекторов: они превращают сигналы универсального для системы кода в специфическую энергию воздействия. Например, сокращение мышцы или выделение слюны. Рецепторы и эффекторы имеют статические и динамические характеристики, подобные нейронным (см. рис. 8, 9). Как правило, активность рецептора продолжается до тех пор, пока на него падает внешняя энергия.

Если представить себе, что каждый рецептор воспринимает энергию из ограниченного пространства внешнего мира или «тела», что он активизирует один элемент в моделирующей установке (например, один нейрон «рецепторного поля» мозга), то совокупность активированных в данное время элементов составит мгновенную модель пространства внешнего мира, или «тела», энергию которого воспринимает система рецепторов (рис. 10). Модель будет изменяться так же быстро, как изменяются воспринимаемые объекты внешнего мира, и с той скоростью, с какой их изменения улавливаются рецепторами. Для превращения такой сугубо временной модели в постоянную нужно, чтобы между активированными элементами установились связи (рис. 11). В этом случае повторное возбуждение извне некоторого числа элементов модели по таким связям могло бы активировать всю модель, иными словами, произошло бы вспоминание воспринятой ранее картины. Однако при этом придется отключать рецепторы, чтобы не наслоились новые картины, воспринимаемые извне. Для непрерывного запоминания новых картин нужно подключать к рецепторам новые поля элементов. Тогда мы получим серию картин-моделей, как на киноплёнке. Невозможно на одни и те же

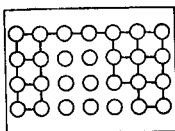


Рис. 11. Схема длительной памяти. Модель представлена проторенными связями между элементами (нейронами).

элементы воспринимать новые картины и одновременно запоминать их.

Интеллект — это целенаправленные действия с моделями, так можно перефразировать первоначальное его определение. Естественно предположить, что действия будут тем более эффективными, чем более подробными будут модели. Это значит: необходимо точное отражение структуры объекта управления в моделях, отражение его изменений во времени и пространстве. Для этого нужно прежде всего множество рецепторов, поскольку объект можно отразить, либо воспринимая его большим количеством точек одновременно, либо сканируя его с высокой скоростью и передавая энергию всякий раз на новые элементы «рецепторного поля». Первый способ используется глазом, второй — телекамерой. Изображение на телевизионном экране светится, то есть запоминается в виде первичной модели, очень недолго. Чтобы запомнить «впрок», следует записать его на видеомagnитофон, а для этого нужно много кадров. Человек воспринимает глазами и на короткое время отпечатывает в мозге огромное количество картин, в которых полезная информация составляет ничтожную часть. Именно ее нужно выбрать, а все остальное — забыть, чтобы не перегружать память.

Обобщенность восприятия

Основная проблема моделирования сложных объектов для целей интеллекта — это преодоление избыточного разнообразия, борьба с избыточной информацией. Чем более развит интеллект, тем более подробными должны быть модели управляемых объектов, тем большие отрезки времени они должны отражать. И тем не менее значительная часть информации о внешней среде, не представляющая ценности для целей интеллекта, должна быть отброшена, модели стерты. Но как произвести отбор? Для этого нужны

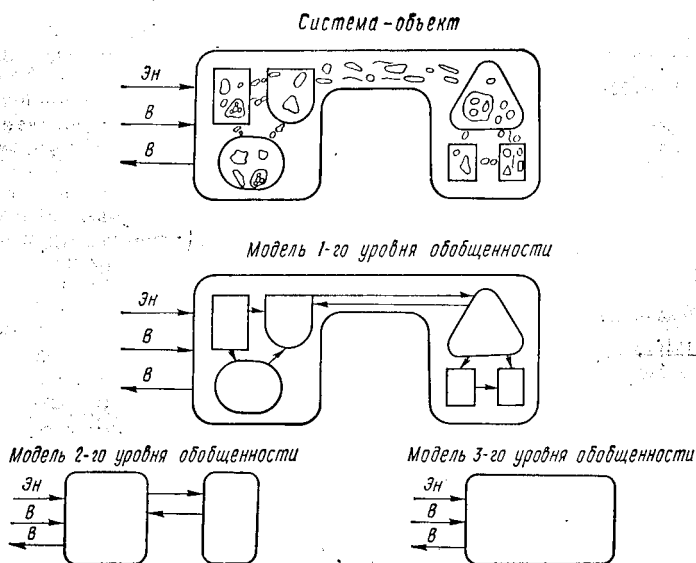


Рис. 12. Объект и его отражение в моделях разного уровня обобщенности.

«Входы» и «выходы»: Эн — энергия; В — вещество.

критерии и способы отбора. При отборе используются все этапы действий с моделями: первичное отражение внешней среды при восприятии, первичное запоминание воспринятого, последующий анализ отраженного в первичной модели. Выход из противоречия между необходимостью запоминать большое количество информации и ограниченностью памяти только один — в огрублении моделей и в отборе их наиболее важных частей. На рис. 12 показано, как можно отразить сложный объект моделями разного уровня обобщенности. Разумеется, по мере упрощения теряется информация, возможно, ценная. Но всегда ли ценность — в подробностях картины, в деталях структуры? Нет, не всегда. Чаще главная информация представлена в грубой модели. И не все детали одинаково ценны. Следовательно, необходимо воспринимать, запоминать и забывать объекты «выборочно», обязательно «в целом» и, кроме того, — интересующие нас детали.

Огрубление модели я называю термином «обобщение». Он относится одинаково к описанию как пространственных, так и временных изменений структуры.

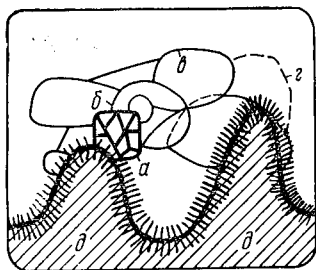


Рис. 13. Модель, отражающая пространственное расположение объектов внешней среды с различной степенью активности и фокусирования. Объект *а* отражен четко и с высокой активностью. Объект *б* находится позади него и отражен менее четко и с меньшей активностью. Еще менее активным является отражение объектов *в* и *г*. Объект *д* находится впереди всех и отражен в модели весьма нечетко.

Понятия «обобщение», «обобщенность» и «степень обобщенности» модели я считаю важнейшими в теории интеллекта. В структурном выражении обобщенная модель — это такая модель, в которой представлены только крупные блоки структуры объекта или пространственного расположения объектов, без деталей (см. рис. 12).

Интенсивность возбуждения рецептора зависит не только от количества падающей на него энергии, но и от изменения его чувствительности, что описывается статической и динамической характеристиками. Яркость восприятия отражается активностью элементов первичной модели во временной памяти «рецепторного поля». Чувствительность рецептора определяется его настройкой и напряжением, регулируемых интеллектом. Восприятие объекта рецепторами непосредственно связано с характером первичной модели (ПМ на рис. 1, 2). «Орган зрения» любого интеллекта должен удовлетворять ряду требований. Первое — должно быть возможным фокусирование зрения, позволяющее ярко, с высокой активностью запечатлевать избранные части (детали) картины, а все остальное представлять в крупных блоках и с меньшей активностью. Второе требование относится к восприятию трехмерных структур. В модели нужно отразить не только часть, выделенную на плоскости, но и часть, выделенную по глубине. Все другие части, расположенные с боков, впереди и позади избранной, необходимо представлять в «крупноблочном» изображении и с меньшей активностью (рис. 13). Настройка рецептора должна обеспечить передвижение фокуса с одной части объекта на другую как по плоскости, так и по глубине, фиксируя эти движения в специальной модели настройки. В результате получится несколько кадров с

плоскими моделями по типу системы моделей, представленной на рис. 17, причем каждый будет маркирован «знаком настройки». Передвижение фокуса или центра восприятия по деталям объекта может представить самостоятельную модель, отражающую его структуру (вспомним детские рисунки из палочек!).

Глаза человека наилучшим образом приспособлены к восприятию внешнего мира с разной степенью обобщенности и разной активностью частей картины. При напряжении внимания предмет рассматривается сфокусированным зрением, при этом мы четко видим мелкие детали в ограниченном пространстве, периферия же видна очень расплывчато, только в крупных структурных блоках. Если намеренно не фокусировать зрение, можно видеть одновременно большое количество объектов, но все они будут нечеткими, расплывчатыми. Разная настройка на глубину позволяет получать модели пространственного расположения объектов с выделением значимой фигуры на фоне второстепенных (см. рис. 13). Угол схождения глаз при сфокусированном рассмотрении дает расстояние предмета от субъекта.

Виды и механизмы памяти

Понятие модели неотделимо от структуры памяти так же, как и от механизмов восприятия. На рис. 1, 2 были показаны два варианта ИИ. Их основное отличие — в носителях памяти, которыми определяются различия в действиях с моделями.

Слово «память» имеет два значения. С одной стороны, это явление, феномен фиксации модели в результате восприятия объекта рецепторами. С другой — это сами запечатленные модели. В последнем случае первостепенную роль играет носитель памяти.

Основной параметр всякой памяти — длительность запоминания. Наиболее короткая память у рецепторного элемента: она длится ровно столько времени, сколько необходимо для накопления энергии, нужной, чтобы выдать в мозг один импульс. В этот момент рецептор освобождается для нового восприятия энергии, его память мгновенная.

На рис. 1, 2 выделен блок запоминания первичной модели *ПМ* — на время, пока она распознается и по ней активизируется распознанная модель *РМ*, которая

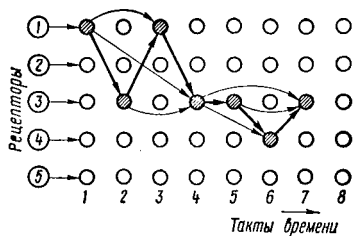


Рис. 14. Схема запоминания последовательности звуков. Возбужденные нейроны «рецепторного поля» заштрихованы.

в свою очередь используется для выбора моделей действия МД. Память для всех этих моделей естественно назвать активной, кратковременной или оперативной в противоположность длительной памяти — основному хранилищу моделей. Возможен и третий вид памяти — «внешняя», находящаяся вне «мозга», вне интеллекта (например, собственные записи, рисунки, которые можно повторно привлекать к использованию). Деление памяти на такие виды условно, но необходимо. В СИ активная и длительная память совмещена на одних сетевых элементах, в АИ массивы памяти совершенно различны.

Характер первичной модели, отражающей трехмерную структуру, был показан на рис. 13. В моделях из длительной памяти структура фона очень неясная и обобщенная, выделяется только фигура. Для алгоритма важны параметры модели и ее частей. Они следующие: активность определяется яркостью объекта, настройкой и напряжением рецептора, четкость — фокусированием рецептора. Расстояние до рецептора для главной фигуры отмечается точно, а для «фона» — приблизительно. Степень обобщенности модели и ее деталей зависит от параметров настройки, расстояния до рецептора и помех.

В СИ мгновенная первичная модель остается в кратковременной памяти наряду с распознанной моделью. В АИ она перекодируется цифровым кодом, который должен отразить все параметры каждого объекта и их пространственных отношений. Именно это и представляет самую трудную задачу для моделирования интеллекта.

Модель во временной памяти выступает как единое целое, следовательно, между ее возбужденными элементами сразу же должны устанавливаться связи, хотя бы тоже временные. Для примера на рис. 14 пока-

зана схема запоминания последовательности из семи звуков, составляющих слово. Память на слова есть у высших животных и птиц. По горизонтали отмечены номера условных тактов времени, по вертикали — номера однородных элементов, последовательно соединяемых с одним и тем же рецептором в каждый такт времени. В первый такт возбужден рецептор 1, во второй такт — рецептор 3, в третий такт — снова рецептор 1 и т. д. Модель станет действовать как единое целое только в том случае, если последовательно активируемые элементы (на рисунке они заштрихованы) в течение семи тактов будут соединены связями и сохранят активность все вместе по крайней мере на время от первого до седьмого такта. Без такого условия — это лишь разрозненные точки, не составляющие единой модели как действующей единицы информации. Отсюда, однако, следует, что каждый рецепторный элемент должен иметь связь со многими элементами «рецепторного поля» и соединяться с ними последовательно, всякий раз с новыми. Это не вызывает особых затруднений для случая, рассмотренного в нашем примере, где один звук соответствует одному элементу в каждый такт. А как обстоит дело, если мы воспринимаем не последовательность звуков, а, скажем, зрительную картину, то есть множество взаимосвязанных объектов? Тут нужны уже не столбцы элементов, а «кадры», которые будут мысленно прокручиваться наподобие киноленты. И сколько же связей нужно установить между возбужденными (яркими) точками кадров!

К сожалению, что-нибудь другое предположить трудно. Можно представить, что энергия с рецептора подается все время на один и тот же, «свой» элемент временной памяти. Допустим, что в следующий момент активируется другой элемент от другого рецептора и между ними устанавливается связь. К следующему моменту первый и второй элементы уже утратили активность, но связь «запомнилась» и т. д. Предположение еще менее вероятное, потому что в таком случае нужны специальные механизмы памяти, отмечающие последовательность возбуждения элементов, чтобы, к примеру, иметь возможность повторить слово. Но для такого «счетчика адресов» тоже нужны структуры.

45 Память человека действительно обширна. Есть люди, которые могут прочесть наизусть поэму «Евгений

Онегин». Но мало таких, которые способны повторить в точности длинное стихотворение после одного прослушивания. Это значит, что «кадры» временной памяти постепенно освобождаются и, таким образом, становятся способными снова запоминать новые модели. Правда, они освобождаются не совсем, кое-что переходит в постоянную память, но очень небольшое, если сравнить объем информации, который мы воспринимаем ежедневно, и то, что из него остается в постоянной памяти. Как правило, мы запоминаем из воспринятого только то, что потом многократно мысленно повторяем.

Гипотеза о механизмах памяти

Трудно предложить гипотезу о механизмах памяти, пригодную для алгоритма интеллекта. Попробую изложить свою попытку. Для обозначения моделей и их сочетаний я буду пользоваться следующими заимствованными из лингвистики терминами. «Буква» — элементарная модель, самый малый значимый признак. «Алфавит» — совокупность букв-моделей, формируемых одним типом рецепторов; «алфавитов» может быть много: световые, звуковые и др. (см. ниже) «Слово» — более сложная модель, состоящая из элементарных моделей, то есть «букв», но очень хорошо организованная и выступающая как одно целое. «Фраза» — соединение «слов», чаще всего временное и непрочное. «Буква обобщенности» — знак, указывающий уровень обобщения моделей, их место в «иерархии блочности».

Модели и ансамбли нейронов. У человека в коре очень много нейронов, их количество оценивают в 10 и более миллиардов. Каждый нейрон соединяется с многими сотнями других. Таким образом, имеются почти беспредельные возможности для образования структур из нейронов, объединенных проходимыми («проторенными») связями. Такие структуры — нейронные ансамбли — могут выступать в качестве моделей объектов внешнего мира. Один нейрон может включаться в несколько ансамблей, как это показано на рис. 15. Модель повторно возбуждается («вспоминается»), если возбудится некоторый процент входящих в ее ансамбль нейронов. Почему не возбуждаются все

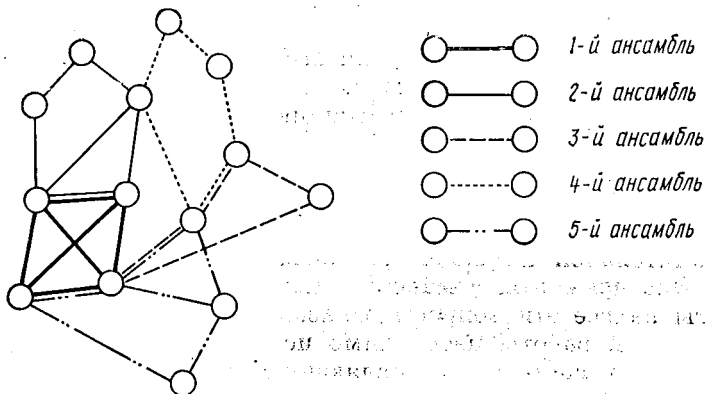


Рис. 15. Нейронные ансамбли.

нейроны коры? Ведь связей между ними вполне достаточно. Чтобы ответить на этот вопрос, нужно сделать еще одно предположение: когда один нейрон возбуждается, соседние тормозятся. Этот принцип «индуктивного торможения» установлен И. П. Павловым. Таким образом, если допустить, что 30% нейронов данной модели возбуждятся извне, то все окружающие должны бы затормозиться. Так и происходит, но связи к остальным 70% нейронов данной модели настолько проторены, что эти нейроны больше возбуждаются по связям, чем тормозятся «по площади». В результате активируется весь ансамбль, вся модель, а другие модели остаются заторможенными, несмотря на то что часть возбужденных в данной модели нейронов входит и в них. Такова гипотеза об ансамблях нейронов.

Для проверки этой гипотезы сотрудниками отдела биокibernетики разработан макет нейронной сети, содержащий 22 узла и наборное поле, которое позволяет осуществлять все возможные соединения между узлами. Узлы (модели нейронов) представляют собой усилители постоянного тока с нелинейной (S-образной) характеристикой. В качестве связей использованы постоянные резисторы. Вес связи считается обратно пропорциональным сопротивлению резистора.

Макет создавался для получения ответа на следующие основные вопросы:

— можно ли технически реализовать устойчиво работающую сеть из нейронных ансамблей так, чтобы

при возбуждении одного ансамбля не возбуждались одновременно и другие, связанные с ним;
— можно ли построить устойчивую сеть, если ансамбли пересекаются, то есть одни и те же узлы входят в состав различных ансамблей;
— можно ли добиться четкого перехода возбуждения с одного ансамбля на другой, если они пересекаются значительной частью своих узлов.

Эксперименты с макетом дали положительные ответы на все эти вопросы. Оказалось, однако, что для успешной работы необходимо использовать централизованную систему, управляющую активностью всей сети. По своим свойствам и назначению эта система аналогична системе усиления-торможения (СУТ), которая подробно будет описана ниже. Следует отметить, что многие исследователи, пытавшиеся построить устойчивую ансамблевую сеть без централизованной системы управления активностью, встретились с большими трудностями и вынуждены были накладывать жесткие ограничения на допустимую структуру сети.

На макете было также показано, что устойчиво работающая ансамблевая сеть может иметь число ансамблей, превышающее число узлов сети. Поэтому в некоторых задачах, используя ансамблевые сети, можно получить более экономную аппаратную реализацию. С увеличением числа узлов эта экономия становится все более ощутимой. К сожалению, построение ансамблевых сетей большого объема связано пока со значительными трудностями.

Устройство «рецепторного поля». На рис. 16 показана гипотетическая схема «рецепторного поля», объясняющая принцип совмещения временной и постоянной памяти. Такое совмещение, несомненно, имеет место в коре мозга.

Система рецепторов R_{ij} подает активность на «рецепторное поле», состоящее из довольно большого количества кадров, построенных так, что в каждом из них имеются представления — элементы для каждого рецептора. Для удобства понимания кадры расположены в виде кольца. Число кадров значительное, но не бесконечное. Предположим, что есть переключатель в центре кольца, который или поворачивает его на один кадр в каждый такт времени, или поочередно переключает связи от рецепторов с одного кадра на другой, соседний. Имеется настройка кадров H , обеспечивающая фокусирование, то есть позволяющая им

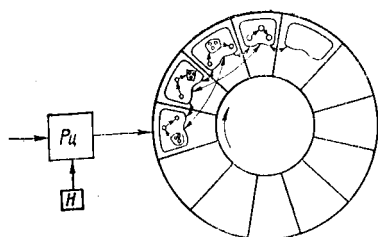


Рис. 16. Схема «рецепторного поля»:

$P_{\text{ц}}$ — рецепторы, H — настройка.

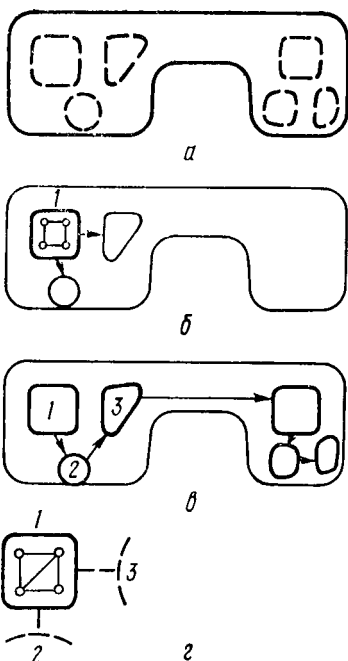


Рис. 17. Система моделей разной степени обобщенности, отражающая объект.

четко воспринимать избранную деталь и неясно видеть весь объект. Она же передвигает фокус по структуре объекта. Еще одно условие: настройка «устает», поэтому после нескольких тактов рецептор отключается совсем.

Характеристики затухания активности элементов очередного кадра после отключения от них рецептора имеют вид, показанный на рис. 8. Затухание активности в самом первом кадре происходит раньше, чем завершится полный круг переключения.

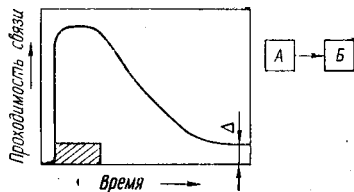
Картины, представленные в кадрах, показаны на рис. 17. В первый такт времени система рецепторов при отсутствии фокусировки неясно воспринимает объект как целое, и он отражается в виде наиболее обобщенной модели (рис. 17, а). Между активированными элементами возникают связи. Предположим, что объект «заинтересовал» интеллект (об этом — ниже), тогда в следующий такт, приходящийся на следующий кадр, зрением, умеренно сфокусированным на верхнем левом углу, преимущественно воспринимается часть

объекта, обозначенная цифрой 1. Создается временная модель этой части, представляющая собой не очень четкое воспроизведение последней при обобщении всего остального объекта (рис. 17, б). Затем фокусировка переходит на часть 2, далее на часть 3 объекта, и они отпечатываются в модели. Модели в предыдущих кадрах еще сохраняют активность, и от них проторяются связи к следующему кадру. В каждом кадре на соответствующих элементах отмечаются направление и степень фокусировки всей системы рецепторов (назовем ее «глазом»). После того как настройка на крупные части обойдет их и они стечатаются на кадрах, структура их расположения вырисовывается в «модели-схеме» (рис. 17, в). Вся серия картин, отображенных в кадрах, объединена продольными связями. На этом восприятие может закончиться, но может быть продолжено на следующий цикл, состоящий в еще большей фокусировке (напряжении) «глаза», нацеленной на детальное рассматривание каждого блока с отражением тонких подробностей его структуры, «привязанных» к блоку, а через него — и к общей структуре объекта. На рис. 17, г этому случаю соответствует модель, формирующаяся при восприятии части 1 объекта сильно сфокусированным зрением. Весь процесс рассматривания объекта запечатлевается в серии кадров. В целом — это «фраза» изучения предмета, запечатленная в кадрах рецепторного поля.

Вспоминание, обобщение, забывание

В первое время вся система моделей в серии кадров активна, и если отключить рецепторы, то можно заново их просмотреть с начала до конца, как бы повторив процесс реального изучения, повторно активируя запечатленные образы. Так человек и делает, мысленно повторяя только что услышанную фразу или воссоздавая в воображении процесс рассматривания сложного объекта. Это бывает в том случае, когда объект или фраза заинтересовывает наш разум. При таком повторении связи между элементами проторяются и происходит процесс перехода временной памяти в длительную, поэтому картину еще можно вспомнить некоторое время спустя. При многократном вспоминании связи проторяются сильно и объект может запом-

Рис. 18. Типичная динамическая характеристика связи между элементами А и Б. Повышенная проходимость связи остается после прекращения возбуждения элемента А (заштрихованный участок на оси времени). Δ — остаточная проходимость связи, определяющая «вклад» связи в постоянную память.



ниться во всех его деталях. Если же значимость предмета не очень велика и повторения не имеют большой активности, то происходит обобщение, то есть постепенное сокращение модели за счет забывания мало значащих ее частей, которые были неактивны при восприятии или не привлекли интереса во время повторения. Так исчезают из памяти целые кадры.

В конце концов может остаться лишь очень обобщенная модель предмета, однако при этом сохраняется воспоминание о самом факте детального изучения объекта, знание о том, что он был изучен подробно. Видимо, это обобщенная модель самого процесса переключения настройки «глаза».

Образ предмета запечатлевается не в одном кадре памяти, а целой их серии, многократно, хотя и с разными деталями. Это соответствует данным физиологии о том, что удаление какой-либо части затылочной области коры не разрушает определенных участков зрительной картины, а просто обедняет ее всю.

Если объект не имел ценности и картина его не вспоминается повторно, то образ совершенно исчезает из памяти, поскольку для проторения первично возникающих связей необходима повторная активация элементов модели. Но, так или иначе, с течением времени происходит закономерное «освобождение» памяти, потому что запомненная картина-модель всегда значительно упрощена по сравнению с воспринимаемой и, следовательно, не занимает все элементы «рецепторного поля».

При восприятии изменяющейся картины рецепторы повторно подключаются к предмету, в результате получается новая серия моделей-кадров, имеющая связи с первой.

В соответствии с этой гипотезой кратковременная и длительная память реализуется на одних и тех же элементах «нейронной сети» в одном рецепторном

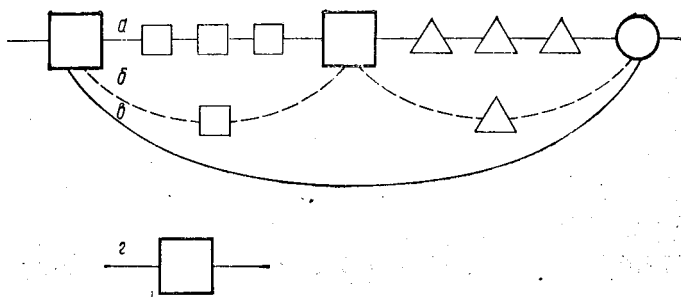


Рис. 19. Схема постепенного забывания и сокращения модели объекта, первоначально состоящей из серии кадров с разной обобщенностью и детальностью, — линия *a*. Менее значимые и похожие детали заменяются одной — линия *б*. Потом остаются только две крайние модели (линия *в*) и, наконец, лишь одна обобщенная модель *г*.

поле, первая — за счет активности элементов, вторая — за счет развития связей. Переход между ними возможен в виде кратковременной памяти связей. Последняя задается характеристикой изменения проходимости связи во времени по такому же типу, как и характеристика изменения активности элемента, но удлиненной во времени и уменьшающейся не до нуля, а до некоторой остаточной величины, определяющей «вклад» связи в длительную память. При повторном использовании данной связи такие «вклады» (Δ) накапливаются и определяют прочность памяти (рис. 18). Если модель повторно не возбуждается, то связи не функционируют, и их проходимость уменьшается. Повторная активация модели сопровождается тренировкой ее элементов, что выражается в изменении ее статической и динамической характеристик, а также в повышении уровня спонтанной, собственной «активности покоя» модели.

На рис. 19 показаны этапы забывания и постепенного сокращения модели, так что в результате остается только несколько обобщенных и связанных с ней «частных» моделей, характеризующих объект лишь в отношении его значимости, то есть полезности.

Разумеется, для того чтобы с помощью элементов одного «рецепторного поля» запоминать все новые и новые модели, нужно допустить наличие большого количества кадров. Кроме того, необходимо предположить торможение моделей как состояние, противопо-

ложное активности, возбуждению. Торможение — отрицательная активность, требующая для своего преодоления дополнительной «мощности», идущей по связи от внешнего источника — рецептора или соседнего возбужденного элемента.

О реализации гипотезы

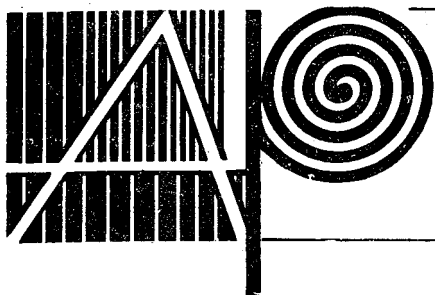
Наша гипотеза предполагает строение «рецепторного поля» в виде сети из элементов с неограниченно большим количеством связей. Примерно такая структура имеет место в коре мозга. Воспроизвести ее техническими средствами пока не представляется возможным, разве что в очень ограниченных пределах, которые едва ли смогут обеспечить демонстративность устройства. Все надежды на алгоритмический интеллект.

В АИ все виды памяти должны существовать раздельно. Время нужно делить на такты и все расчеты активности моделей и проходимости связей осуществлять «ступенчато», от такта к такту. Первый вид памяти в АИ — это картина с рецептора. Она существует очень короткое время и считывается, перекодируется по определенным правилам, которые еще нужно создать. Получается ряд цифр, отражающих как саму структуру объекта, так и перечисленные выше параметры модели. Главный из них — это уровень активности каждого объекта модели картины. Поскольку процесс рассматривания даже неподвижной картины выражается в серии кадров типа показанных на рис. 17, то и цифровое выражение модели объекта будет состоять из нескольких строк цифр, кодирующих каждый кадр восприятия. Связи выразятся адресами кадров. Тогда два-три кадра составят «фразу» — модель объекта во временной памяти. Параметр активности кадра в целом и его отдельных объектов будет понижаться по определенной характеристике, сходной с представленной на рис. 18. По мере отдаления во времени кадры станут «бледнеть», утрачивать детали, так что вся «фраза» будет становиться все короче и короче. Если не произойдет повторного привлечения внимания и активации «фразы», модель сотрется из памяти. Если же активность и связи будут подновляться повторным использованием модели, то через некоторое время, предусмотренное характеристикой, модель перейдет в

длительную память вместе со своими связями-адресами. Пересчет активности и связей всех моделей в кратковременной памяти обязателен для каждого такта.

Длительная память в АИ выражена «фразами», перешедшими из кратковременной памяти. Сейчас трудно представить всю организацию массива памяти. Думаю, что он должен состоять из большого числа «словарей фраз», построенных из 2—4 «слов» каждая. Во «фразах» будут широко использоваться обобщенные модели.

В СИ все модели постоянно сохраняют хотя бы минимальную активность, вследствие чего связи между ними постоянно изменяются. Это очень затрудняет воспроизведение СИ на ЦВМ, поскольку с увеличением объема сетей катастрофически возрастает объем расчетов. АИ позволяет уменьшить расчеты за счет удлинения интервалов времени между пересчетами связей массива длительной памяти.



Действия с моделями

Здесь мы рассмотрим только важнейшие действия с моделями. К ним можно отнести действия активации моделей, их сравнения, а также дописывания «фразы» и обобщения моделей.

Активация моделей

В памяти находится масса моделей, составленных из «слов», «фраз», «букв» разных «алфавитов». Модели объединены связями, по которым они взаимодействуют друг с другом. Большинство моделей находится в неактивном состоянии. В частности, это касается всех моделей длительной памяти АИ и в меньшей степени — СИ, в котором нет разделения активной и пассивной (кратковременной и длительной) памяти. Деятельность интеллекта связана с активацией новых моделей в длительной памяти и постепенным затуханием активности моделей в кратковременной памяти. В мозге и в ИИ на физических сетях каждый элемент модели — нейрон — или целую модель — ансамбль из нейронов — можно представить как генератор специальной («нервной») энергии, возникающей в ответ на действие такой же энергии, которая поступает по связям от других моделей. Генератор

работает по статическим и динамическим характеристикам, подобным показанным на рис. 8 и 9. Энергия передается по связям на другие модели; количество ее определяется проходимостью связи.

Активное состояние модели можно назвать физиологическим термином «возбуждение». В нейронах мозга оно выражается частотой импульсов, в СИ на физических сетях — это электрический потенциал. В ИИ, моделируемом на цифровых машинах, уровень активности моделей — это главный параметр, «буква», выраженная числом, и его нужно пересчитывать для каждого временного такта по статическим и динамическим характеристикам. Впрочем, для АИ это касается только моделей в кратковременной памяти. Операции активирования моделей могут быть двух видов: извлечение модели из длительной памяти с расчетом ее активности или пересчет уровня активности модели, уже находящейся в кратковременной памяти, если она получает дополнительный импульс по связям от другой модели.

В СИ выбор новой модели для активации определяется структурой связей, идущих от активной модели. В АИ новая модель вызывается из длительной памяти по «адресу», записанному в «словаре фраз», в котором первым «словом» является уже возбужденная модель. Например, есть «словарь» предмет—действие, в нем есть модель «хлеб», ей соответствует модель действия «жевать». Последняя и будет вызвана, если в оперативной (кратковременной) памяти содержится возбужденное «слово» «хлеб». Уровень активности модели «жевать» будет подсчитан, исходя из статической характеристики коэффициента проходимости связи, записанного в «словаре», и активности модели «хлеб».

В соответствии с нашей гипотезой для функционирования интеллекта необходимо еще другое состояние, противоположное по знаку возбуждению, — так называемое «торможение». Этот термин принят в нейрофизиологии. Мы его представляем как отрицательную активность, которая тоже генерируется специальными центрами и вычитается из положительной активности при расчетах. Впрочем, необходимость в торможении нужно еще уточнять при проектировании ИИ.

Сравнение моделей и распознавание образов

Второй тип операции с моделями — это их сравнение между собой с целью установления как общности, так и различия. Реализация действия целиком зависит от вида интеллекта и организации памяти. В мозге сравнение осуществляется, по всей вероятности, путем условного «наложения» моделей друг на друга. При этом их сходство и различие определяются по количеству общих элементов. Из физиологии известно, что очаг возбуждения в коре генерирует торможение на окружающие участки, затем возбуждение первого очага падает, его «соседи» освобождаются от торможения, и возбуждается другой очаг коры. Можно предполагать, что этим следующим очагом, то есть моделью, будет такой очаг, у которого много связей с первым или много общих нейронов в составляющих модели ансамблях. Сходство и различие определяются по отношению к каким-то третьим моделям-признакам, связи к которым идут от первой и второй из сравниваемых моделей. Допустим, что первая возбужденная модель вызвала к активности признак 1, а вторая — активированная по сходству — возбудила признак 2. Степень совпадения признаков — это мера общности и различия моделей. У животных нет количественного выражения для этой меры, у человека, овладевшего счетом, она есть.

Для АИ сравнение моделей — банальная операция вычитания двух строк цифр. Выраженная цифрами модель разделена на разряды со своими значениями. Можно предположить, что в первом разряде представлена наиболее обобщенная модель (какое-то материальное тело), во втором — крупные структурные блоки (голова, туловище, ноги, отличающие человека), в последующих разрядах — детали. Такой образ всегда имеет место, когда мы воспринимаем объект, даже при фокусировке зрения на его деталях. По этим разрядам и будет осуществляться сравнение.

Сравнение известной модели с неизвестными лежит в основе распознавания образов. По модели объекта, отпечатанного с рецептора в кратковременной памяти, которая не имеет связей с другими моделями и, сле-

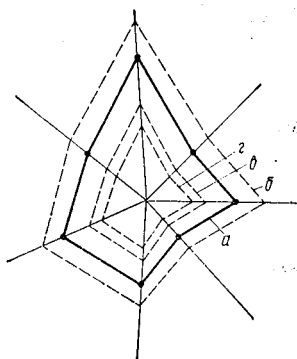
довательно, является неизвестной, нужно найти модель-эталон, имеющую такие связи, иначе говоря, входящую в различные «фразы» и числящуюся в «словарях». Именно связями определяется то, насколько знаком нам тот или иной объект: чем больше связей, тем лучше мы его знаем. Вероятность распознавания определяется точностью совпадений неизвестной модели с эталонами. Множественное число я употребил не случайно: объект может походить на несколько других, известных.

Распознавание в СИ осуществляется автоматически: ансамбль возбужденных с рецептора элементов, который представляет собой модель неизвестного объекта, накладывается на другую модель. Она активируется, а затем активируются связанные с ней модели, опознающие объект. Поочередно может активироваться несколько похожих моделей, каждая со своей степенью сходства.

В АИ для распознавания модели нужна специальная программа извлечения из постоянной памяти серии моделей и сравнение каждой из них с моделью объекта. Выборка моделей из памяти должна производиться начиная с самого обобщенного признака — «буквы». По ней выбирается «словарь» и далее сравниваются вторые и следующие «буквы», так же как производится поиск значения «слова» по «словарю». «Известность» наиболее близкого из искомых «слов» определяется числом вхождений его в «словари фраз». Степень вероятности опознания объекта определяется совпадением последних «букв» — деталей, потому что по первым «буквам», определяющим обобщенные признаки, всегда можно найти много похожих. Человека легко отличить от других объектов, труднее распознать — кто есть кто.

Остановлюсь на двух обстоятельствах, осложняющих распознавание. Первое — «неполнота» модели объекта, обусловленная помехами восприятия, дальностью расстояния или недостаточным напряжением рецептора. Неполнота или неясность первичной модели выражается в отсутствии ряда деталей, в «крупноблочности». При этом всегда присутствует «буква», объясняющая неполноту, — показатель низкой настройки рецептора или наличия внешних помех. Я намеренно не употребил понятие «обобщенность» применительно к такой модели, потому что оно предусматривает выражение модели крупными блоками в результате спе-

Рис. 20. Схема гипотетических «рельсов» в «рецепторном поле», позволяющих производить приведение модели к одному определенному размеру. В памяти хранится модель a . При восприятии объекта с близкого расстояния большая модель b уменьшается до размеров a ; при восприятии объекта с большого расстояния малые модели c или z увеличиваются до размеров a .



циального отказа от деталей, а не отсутствия их из-за плохого восприятия. Неполную модель можно распознать, только сравнивая ее с обобщенными моделями-эталоны, чем и определяется полнота распознавания. Например, видно, что объект — человек, но мужчина это или женщина, определить нельзя из-за неясности образа. Более четкую первичную модель можно получить за счет настройки рецепторов или приближения к объекту.

Второе обстоятельство — это различие в размерах первичной модели и моделей-эталонов. Общеизвестно, что человек может распознать объект с разного расстояния, если он хорошо изучен вблизи. Распознавания прямым наложением моделей здесь не получится. Нужно допустить специальный механизм приведения модели к одному определенному размеру в виде своеобразных «рельсов» в «рецепторном поле», как показано на рис. 20. «Рельсы» эти позволяют изменять размер первичной модели, сохраняя сходство. По всей вероятности, нечто подобное есть в зрительной области коры. Для АИ перекодирование первичной модели цифровым кодом должно предусматривать приведение к стандартному размеру моделей-эталонов.

Есть еще ряд обстоятельств, затрудняющих распознавание: различия исходных положений объекта, его деформации и др. Многие из возникающих здесь вопросов подробно исследовались кибернетиками, и полученные ими результаты можно применить и для АИ.

Дописывание «фразы» — вспоминание

Операцией дописывания «фразы» можно назвать активацию какой-либо одной модели, являющейся следствием активации другой модели. Активация происходит по связи, соединяющей обе эти модели. Такую операцию можно еще обозначить термином «вспоминание». Активация последовательности «слов» во «фразе» осуществляется по принципам, описанным в начале этого раздела. «Фразы» могут быть самыми различными, в них может запечатлеваться любая последовательность воспринятых образов, закрепленная в памяти повторным вспоминанием. В АИ «фразы» записаны в «словарях фраз». Их много: предмет—действие, предмет—качество и др., и наоборот. Важными являются, так сказать, вертикальные «фразы»: «вверх» — от детали к обобщению, «вниз» — как расшифровка обобщенной модели детальными вариантами. Память в АИ должна состоять из коротких «фраз» в 2—4 «слова», а длинные последовательности моделей должны состояться из нескольких «фраз». Это проще, чем создавать «словари» длинных «фраз». Человеческий разум тоже оперирует короткими «фразами» образов. Вызов следующего «слова» «фразы» возможен только в случае достаточной активности первого «слова» и достаточной проходимости связи от него ко второму. Активность рассчитывается по входам на первое «слово» от других моделей (по другому «словарю»), а параметр связи записан в «словаре».

Обобщение моделей

Операция выделения обобщенной модели из серии конкретных моделей осуществляется их последовательным сравнением и выделением общего признака сходства между ними (пример — «четвероногие»). При этом все остальные признаки, по которым модели серии отличаются друг от друга, становятся все более неясными, иными словами, в обобщенной модели они выступают с низкой активностью. Такая обобщенная модель, полученная на материале ряда конкретных моделей, имеет «букву обобщения», ука-

зывающую название действия, в результате которого модель образовалась. Этой «буквой» она отличается от неясной модели конкретного объекта, при которой стоит «буква восприятия».

Замена ряда конкретных моделей одной обобщенной является весьма распространенной операцией. Идя по лесу, вы видите множество деревьев. Все они отражаются во временной памяти, и некоторое время спустя еще можно припомнить отдельные деревья. Однако потом конкретные образы заменяются неким обобщенным деревом, усредняющим виденные распространенные экземпляры («лес из высоких сосен»). При этом обобщенная модель сопровождается «буквами» с адресами действий, отрезков пространства или времени. Обычно обобщенная модель имеет структурный ранг на одну степень выше, чем конкретные составляющие. Например, в «обобщенном дереве» ветки и листья не дифференцируются по породам. «Обобщенный человек» выглядит бесполом, поскольку его черты совершенно неясны.

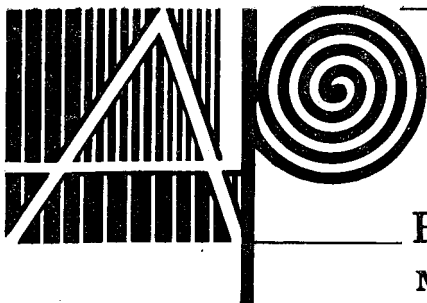
Обобщение касается всех видов моделей — зрительных и слуховых образов, моделей действий, моделей качеств, иногда самых специфичных. «Ранг обобщения» может быть самым различным, но образное выражение обобщений высокого ранга затруднительно, и они выражаются только с помощью речи. Например, можно представить «обобщенный стул» в виде неясного образа предмета со спинкой, сиденьем и ножками, но как зрительно представить себе «обобщенную мебель»? Здесь совершенно разные предметы обобщены по признаку функции или места нахождения. Если потребовать: «Представьте мебель», то выступает ряд неясных предметов мебели и «буква обобщения», указывающая, что есть еще много подобных образов. Только речь позволила обозначить этот уровень обобщения и таким образом дала возможность произвести сами действия.

Можно ли образно представить «вещь»? Или «материальное тело»? Можно, в виде неясной структуры, отграниченной от других. Зрительный образ такой структуры имеется как «что-то», но выделить его позволили только слова речи — «вещь», «тело», «существо», «человек», «животное». Можно предполагать, что черты обобщения всегда присутствуют в конкретной модели. Так, например, понятие «материальное тело» присутствует в модели стула, отражая то обсто-

ительство, что стул — это пространственная структура, ограниченная от среды.

Для цифрового кодирования образов при создании АИ нужно, вероятно, отразить отдельными цифрами принадлежность данной конкретной вещи к основным уровням структурной иерархии обобщения. Для этого необходимо создать систему обозначений. Разумеется, нельзя в модели представить все возможные параметры, по которым можно обобщить, например, стулья, но следует отметить, что они принадлежат к приспособлениям для сидения, к мебели, к предметам (то есть неживым телам). Это удлинит конкретную модель, но упростит операции с ней. Впрочем, не стоит вдаваться в детали, поскольку возможны разные системы перекодирования образов в цифровые модели.

На этом я закончу рассмотрение операций с моделями, хотя далеко не исчерпал их варианты. Важно усвоить главные, потому что они постоянно присутствуют в программах интеллекта, и хотелось бы при дальнейшем изложении не касаться деталей операции, а просто обозначать ее.



Взаимодействие моделей

В ИНТЕЛЛЕКТЕ

В предыдущих главах мы разобрали следующие вопросы:

1. Принципы структурного и энергетического построения среды, в которой действуют интеллекты. Два основных типа искусственного интеллекта — сетевой (СИ) и алгоритмический (АИ).

2. Понятие о моделях, их структурное и энергетическое выражение. Хранение моделей — память.

3. Основные программы действий с моделями — в смысле их превращений и обращения с разными видами памяти.

Последние два пункта разделены довольно условно.

Теперь можно перейти к главному: как должны взаимодействовать модели в интеллекте, чтобы в результате выполнялось его назначение — управление объектами внешней среды, собственным «телом» и даже самим собой.

Если возможно бесконечное количество моделей любого сложного объекта, то мыслимо и бесконечное количество интеллектов для управления им. Поскольку трудно сразу объяснить самые высшие проявления интеллекта, придется проследить его эволюцию от простого к наиболее сложному.

Объекты внешней среды воспринимаются рецепторами, которые направляют соответствующий код в моделирующую установку («мозг»). Здесь создаются

модели среды, по которым выбираются модели действий, реализуемых через органы действий или эффекторы (у живых существ — мышцы). Кроме того, интеллекту должно быть обязательно придана энергетическая установка («тело»), которая снабжает энергией мозг, рецепторы и особенно — эффекторы, поскольку они должны развивать некоторые усилия для воздействия на объект. В свою очередь «тело» должно получать энергию или вещество извне, причем в результате деятельности интеллекта, направленной на среду. Важнейшим элементом интеллекта являются критерии управления объектом, которые в то же время управляют действиями с моделями. Получается своеобразная цепь: объект управляется через эффекторы, последние — моделями, а сами модели управляются критериями. Какова же природа критериев?

Один из критериев ясен: для того, чтобы интеллект мог выполнять свои функции, он должен получать энергию, следовательно, должен обеспечивать снабжение «тела» извне, и качество этого снабжения является критерием деятельности интеллекта. Итак, «тело» диктует разуму. Это старая истина, вполне объясняющая разум животных. Он выполняет роль компьютера для реализации программ инстинктов, заложенных в генах. Однако даже для животных дело обстоит не так просто: критерии для своей деятельности формирует сам мозг на основе восприятия и переработки сигналов, поступающих от тела. И еще сложнее: некоторые критерии черпаются из внешней среды. Никакого парадокса тут нет — животное само является частью среды, продуктом эволюции, в которой среда всегда присутствует в качестве фактора отбора.

Разум животных давно превзойден разумом человека. Поэтому понятие критериев гораздо сложнее, чем кажется на первый взгляд.

Как говорилось, при познании сложных систем возможно бесконечное множество моделей. Это касается не только моделирования неизменных структур, пределом которого является повторение, копирование объекта, как делается в технике. Изменение объекта во времени — тоже предмет познания, поскольку само управление состоит в изменении его во времени в нужном направлении. Предметом моделирования становятся прошлое, настоящее, будущее и не только сам объект, но и среда, его окружающая. Так же бесконечны пределы моделирования управляющих действий,

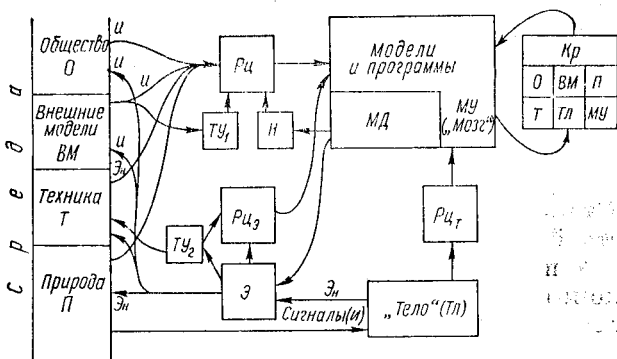


Рис. 21. Расширенная схема интеллекта.

П — природа; *Т* — техника; *ВМ* — внешние модели; *О* — общество. «Входы» и «выходы»: *И* — информация, сигналы; *Эн* — энергия; *ТУ* — технические устройства; *Рц* — рецепторы; *Рц₂* — рецепторы органов воздействия; *Н* — настройка рецепторов; *Рит* — рецепторы «Тела»; *МУ* — моделирующая установка «Мозг»; *МД* — модели действий; *Э* — эфффекторы.

поскольку их можно создавать в процессе творчества, а не ограничиваться выбором из заданных. Более того, так как критерии заложены в мозге и тоже являются моделями некоторых качеств — объекта, среды, тела, самого мозга, то возникает возможность умножать критерии. Разум может быть относительно самостоятельным фактором развития мира.

На рис. 21 показана схема интеллекта уровня человека и его отношения к внешнему миру, к среде. На схеме она представлена четырьмя составляющими. Это прежде всего *природа*, понимаемая во всей совокупности ее факторов. Затем — *техника*, под которой подразумеваются орудия воздействия на природу. Еще выше *внешние модели* — это наука. Наконец, наверху — *общество*. Новая схема отличается от предыдущей тем, что рецепторы получают не только энергию, косвенно сигнализирующую о внешней среде, но и прямую информацию в виде сигналов — от *общества* и *внешних моделей*. Между средой и рецепторами *Рц* встроены технические устройства *ТУ₁*. Это всевозможные приборы, предназначенные для усиления внешних воздействий и превращения невоспринимаемых органами чувств видов энергии в воспринимаемые. Точно так же эфффекторы *Э* могут быть вооружены инструментами *ТУ₂*, значительно увеличивающими мощность их воздействий. От эфффекторов идут стрелки, показывающие движение не только энергии, но и информа-

пии — сигналов, направленных в высшие системы, например в *общество*. Рецепторы имеют органы настройки *H*, а эффекторы — свои собственные рецепторы $P_{ц_0}$, обеспечивающие сигналы обратной связи. Рецепторы есть и в теле ($P_{ц_т}$). Моделирующая установка МУ («Мозг») содержит модели внешней среды, тела, а также модели собственных программ и моделей действий *МД*. Наиболее разнообразны критерии *Кр*: в них представлены не только значимые качества *тела*, но и всех отделов внешней среды и, что самое интересное, — некоторые качества работы МУ («мозга»).

В любом интеллекте заложены исходные модели, первичная структура связей и основные критерии. Все это обеспечивает начало деятельности. Дальнейшее саморазвитие интеллекта зависит от количества элементов и их исходных характеристик — главным образом их тренируемости и способности к образованию связей. Такой интеллект, получая извне информацию, становится способным к обучению и воспитанию под действием общества и его моделей. Воспитание понимается как изменение первичных критериев и формирование новых. Высшим проявлением развития интеллекта являются творчество, самопознание и самовоспитание, то есть способность не только создавать новые модели воздействий и претворять их в вещи с помощью техники, но и формировать новые критерии, изменяющие направление деятельности интеллекта в сторону от программ, заложенных при его создании и в воспитании со стороны общества.

Высшим проявлением интеллекта является воспроизведение самого себя техническими средствами и последующий неограниченный рост его мощности. Многие фантастические предположения в этом направлении вполне допустимы, и их реализация является лишь вопросом времени и средств. Но разговор об этом еще впереди. А пока обратимся к подробному рассмотрению программ интеллекта, начиная с наиболее простых его вариантов.

Критерии, потребности, чувства, стимулы

Основу деятельности интеллекта как действий с моделями составляют изменения активности его элементов. Активность — это энергия. В живых системах каждый нейрон генерирует специфическую

энергию нервных импульсов, источником которой является химическая энергия, получаемая в каждой клетке в ее митохондриях. Они преобразуют энергию глюкозы и жиров в фосфорные соединения, которые и используются клеткой для своих нужд: в нейроне — на нервные импульсы, в мышце — на механическое сокращение, в железе — на синтез новых молекул. Мощности этой «электростанции» изменяются за счет тренированности или детренированности — процессов, которые в свою очередь определяются «запросами на специфическую функцию», предъявляемыми к клетке организмом как целым через его регулирующие воздействия.

В искусственном интеллекте также необходимо сохранить параметры активности элементарной модели, ее тренированности и детренированности как важнейшие средства его деятельности. Без этих параметров я не мыслю возможности создания ИИ. Другое дело, что энергетические затраты на активацию элементов ИИ могут быть ничтожно малы и не иметь значения (впрочем, они невелики и в организме), но выражение активности числом необходимо для осуществления программ взаимодействия моделей.

В живом мозге каждая нервная клетка — тренируемый «генератор импульсов», но для того чтобы она функционировала, выдавала эти импульсы, ей нужна стимуляция, то есть раздражение извне. Источником раздражений является прежде всего тело и во вторую очередь — внешняя среда. Тело действует не только через рецепторы, но и при помощи активных химических веществ, целенаправленно раздражающих центры в подкорке, являющиеся носителями тех самых критериев, о которых говорилось выше. Рецепторы, воспринимающие воздействия внешней среды, не только активируют нейроны «рецепторных полей», но действуют и более сложным образом. Путем обработки моделей среды выделяются сигналы, воздействующие на специфические нейронные ансамбли, которые соответствуют критериям — определенным качествам внешней среды, влияющим на поведение животного наряду с «требованиями тела», такими очевидными, как голод. Примером «внешних» критериев является любознательность. Этот критерий — стимул поведения животных, хотя для него нет «представительства» в теле, а качество новизны, возбуждающее центры любознательности, принадлежит среде.

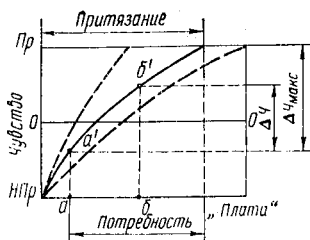


Рис. 22. Зависимость чувства от «платы».

Чувство изменяется от «неприятно» НПр до «приятно» Пр в зависимости от удовлетворения потребности измеряемой величины «платы», которую нужно добавить к ее исходному уровню a , с тем чтобы получить полное удовлетворение. Реальный стимул к некоторому действию для удовлетворения потребности измеряется приращением чувства $\Delta\chi$ от исходного уровня a до уровня b ожидаемой «платы». Притязание — это «плата», необходимая для получения максимума чув-

ства. Пунктирными линиями показаны новые характеристики потребностей, получаемые после адаптации: если потребность систематически не удовлетворяется, характеристика смещается влево, то есть притязание уменьшается, если же потребность удовлетворяется, характеристика смещается вправо, то есть притязание возрастает.

Понятие об информационных критериях как некоторых моделях-центрах, которые возбуждаются путем выделения информации из воздействий среды, но в то же время сами являются источником активности для других отделов мозга, исключительно важное. Его значение еще недостаточно осознано физиологами. У животных есть врожденные модели — программы переработки внешней информации и выделения из нее значимых, ценных качеств — критериев. Их субстрат — «центры» критериев — представляется особенно активным «генератором» нервной энергии, стимулирующим деятельность мозга. Значимые качества внешней среды могут быть самыми различными: новизна, превосходство собратьев по стае и пр.

Структурно критерии можно представлять при помощи моделей, имеющих «входы», «выходы» и характеристики примерно такого типа, как показано на рис. 22. Поскольку критерий является источником энергии для деятельности интеллекта, его можно связать с понятием стимула. Возможен случай, когда в результате деятельности совершаются процессы, дающие отрицательный стимул, то есть тормозящие исходную деятельность.

С понятием критерия тесно связано еще одно психологическое понятие — потребность, неудовлетворение которой побуждает к действию. Потребность является источником энергии для активации моделей. Ее можно выразить через критерий, как это сделано на рис. 22. Логично предположить, что когда потребность полностью удовлетворена, стимул для деятельности, направленной на ее удовлетворение, должен прибли-

зиться к нулю. И наоборот, когда потребность совсем не удовлетворяется, иными словами, когда субъект — интеллект — не получает никакой «платы» от среды, стимул должен быть максимальным. Словом «плата» я буду широко пользоваться, понимая под ним все материальные, информационные и любые другие воздействия среды, получаемые интеллектом в ответ на его деятельность.

Явление адаптации тоже известно в психологии: если животное голодает, оно готово есть любую пищу, если же оно все время сыто, его пищевой центр адаптируется и требования к пище возрастают — подавай только вкусную. Однако если его снова заставить поголодать, все станет на свое место. Особенно распространена адаптация к приятному, к избытку «платы». К сильному голоду адаптироваться нельзя. При чисто психических функциях возможна значительная адаптация и к недостатку «платы» (пример — бедность информации). Уровнем притязаний можно назвать величину «платы», которая необходима для полного удовлетворения потребности, когда стимул приближается к нулю. Выражение конкретной потребности — это приращение «платы» от ее данного среднего уровня до уровня притязаний, до насыщения.

Понятия потребности и стимула тесно связаны с другим термином из психологии: «чувство». Его можно трактовать как субъективную меру удовлетворения потребности. Известно, что разная степень удовлетворения одной и той же потребности дает противоположные чувства — приятные или неприятные. Полный голод, когда потребность не удовлетворена, — неприятен. По мере насыщения наступает момент «нулевого чувства», если продолжать есть, появляется отчетливо приятное чувство. Оно достигает максимума при полной сытости, и стимул к еде в этот момент равен нулю. Стимул как источник деятельности начинается от полного голода, уменьшается по мере насыщения и исчезает при полной сытости. Таким образом, он изменяется от нуля до максимума. Величина чувства изменяется от отрицательного значения (неприятное чувство) до положительного (приятное). Стимул — это приращение алгебраической суммы чувств $\sum \Delta \chi$ от исходного до ожидаемого конечного их состояния.

Психологические компоненты приятного и неприятного наличествуют в каждой потребности. Суммарный уровень душевного комфорта (УДК) или

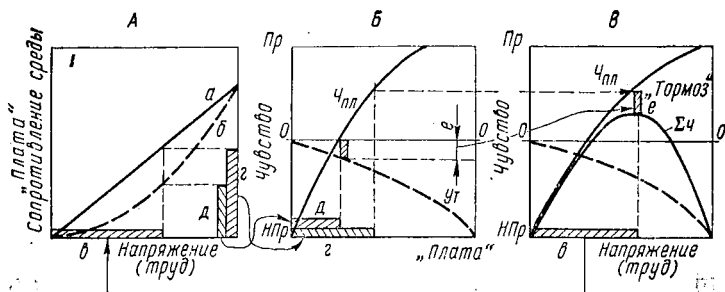


Рис. 23. Формирование оптимального напряжения действия (труда) при взаимодействии системы и среды.

А — характеристики среды: а — изменение величины «платы» в ответ на напряжение; б — изменение величины сопротивления среды в ответ на напряжение. Б — характеристики рабочего элемента: изменения положительного чувства $\chi_{пл}$ в ответ на «плату» и чувства утомления $Ут$ от преодоления сопротивления среды. В — суммарная характеристика: сумма чувств $\Sigma\chi$ имеет максимум, определяющий интенсивность труда. Уровню напряжения в соответствует «плата» г, сопротивление среды д и «тормоз» деятельности е.

усредненный уровень счастья — это тот показатель, к максимуму которого стремится каждый интеллект в живой природе, соответствующим образом выбирая свою деятельность. Этот показатель, как будто не имеющий отношения к искусственному интеллекту, тем не менее и для него является важнейшим рабочим параметром, так же как понятия чувств. Как ни странно, но мне кажется невозможным создание высоко развитого искусственного интеллекта без этих параметров. Конечно, их можно было бы обозначить какими-то абстрактными символами, но это не упростило бы понимания сути дела. Я глубоко убежден, что без моделирования многих, казалось бы, чисто человеческих или даже животных чувств, создание ИИ вообще невозможно.

На рис. 23 показано формирование оптимального напряжения действия (труда) как результат суммирования критерия-стимула и критерия-«тормоза» функциональной системы, взаимодействующей со средой. «Плата» а подается на рабочий элемент, формирующий соответствующее чувство $\chi_{пл}$. Одновременно деятельность встречает сопротивление б, вызывающее утомление $Ут$. Это «тормоз», ослабляющий стимул $\chi_{пл}$. За счет нелинейности обеих характеристик устанавливается некоторое равновесие между насыщением критерия-чувства, «тормозом» и уровнем напряжения деятельности. Практически же происходит колебатель-

ный процесс за счет тренированности и детренированности рабочего элемента. С другой стороны, управляющий элемент — критерий-стимул — действует по другим законам: он способен к адаптации.

Принцип адаптации состоит в том, что характеристика управляющего элемента, в частности генератора критерия-стимула, своеобразно изменяется. Если потребность постоянно удовлетворяется, то крутизна его характеристики уменьшается (притязания возрастают); если она недостаточно удовлетворяется, то она увеличивается (притязания уменьшаются). Статическая характеристика рабочего элемента зависит от тренированности (см. рис. 9). Динамика адаптации и тренированности, то есть скорости изменения характеристики управляющего и рабочего элементов, неодинакова, поэтому саморегулирующийся комплекс, составленный из них, функционирует даже в неизменяющейся среде с некоторыми колебаниями около среднего уровня, на который был «спроектирован» комплекс — природой или конструктором. На рис. 9 показаны характеристики тренированности и детренированности рабочего элемента, а на рис. 22 — адаптация управляющего элемента при разной степени удовлетворения потребности. Такая система рассчитана на работу в среде, сопротивление которой не изменяется.

В действительности же среда не стабильна. «Сопротивление» ее улавливается «тормозом». В первом приближении — это «утомление» рабочего элемента в зависимости от сопротивления его деятельности со стороны «выхода» в среде, на которую направлена его работа. На рис. 23 показаны характеристики «тормоза» в зависимости от напряжения деятельности рабочего элемента и сопротивления среды.

«Плата» — это совсем не обязательно внешние воздействия среды. Источником активности могут быть внутренние структуры интеллекта (мозга), например модели некоторых качеств внешней среды, выделенные в процессе анализа ее моделей. Важно другое: модель, обозначающая критерий-стимул, становится усилителем, генерирующим избыточные мощности, которые этот усилитель передает другим моделям, стимулируя их активность. Проще всего предположить, что подобные модели-усилители запроектированы в генах живых существ. В действительности так и есть для некоторых врожденных чувств, выражающих потребности как функции инстинктов. Однако у

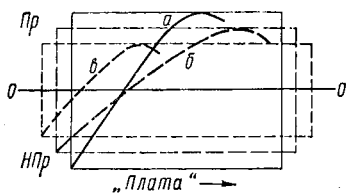


Рис. 24. Характеристики потребностей и соответствующих чувств а, б, в. Разная значимость чувств выражается в различиях абсцисс на универсальной шкале «неприятно» — «приятно» (НПр—Пр).

человека, кроме биологических потребностей, появились социальные, часть которых в генах не предусмотрена и которые выражаются в так называемых «убеждениях». По мощности стимулов они иногда конкурируют с биологическими. Социальные потребности выражаются привитыми обществом моделями словесных формул, например долга («как надо»). Видимо, усилителем может стать любая модель, если ее активно тренировать. Биологический субстрат разума — нервные клетки — это позволяют. Соответственно тренируются и связи с этими «приобретенными чувствами», так что в результате по значимости они действительно становятся в ряд с биологическими чувствами, которые уже изначально заложены как усилители. Между прочим, воспитанием можно детренировать либо натренировать и биологические центры потребностей, и их «коэффициент усиления» изменится.

Можно предположить, что как у моделей биологических, так и у натренированных приобретенных потребностей высок уровень собственной, спонтанной активности, обеспечивающий им постоянное воздействие на другие модели.

Значимость потребностей (или, иначе, их активность, «коэффициент усиления») различна. Модельно это выражается в выборе масштаба шкалы на оси абсцисс на рис. 22, где представлена характеристика «плата» — чувство. Разница в значимости чувств постоянно наблюдается у животных и человека: одни — жадные, другие — агрессивные, третьи — любопытные и т. д. В их различиях выражается направленность интеллекта — система приоритетов его деятельности. Разная значимость потребностей-чувств показана на рис. 24.

Модели потребностей — чувств-стимулов — это главные модели, они всегда находятся в оперативной памяти, поскольку имеют относительно высокую активность, которая периодически еще более возрастает, 72

если потребность долго не удовлетворяется. Активность уменьшается после получения «платы». В этом смысле важны динамические характеристики моделей чувств. Пример статической характеристики показан на рис. 22.

На что расходуется энергия чувств в интеллекте? Можно предположить, что на возбуждение по связям других моделей, например для вызова их из длительной памяти в АИ. (Впрочем, положение о расходовании энергии активной моделью, воздействующей на другую, еще требует проверки расчетами. Возможно, энергия модели просто затухает согласно динамической характеристике, если на «вход» не поступают новые сигналы). Энергия чувств не уменьшается до тех пор, пока соответствующая потребность не удовлетворится получением «платы». Как говорилось выше, все модели в СИ обладают самостоятельной «спонтанной» активностью, величина которой определяется уровнем тренированности модели. Это положение особенно относится к моделям потребностей-чувств. Их постоянная активность должна найти отражение не только в СИ, но и в АИ, поскольку модели потребностей все время находятся в оперативной памяти.

Классификация потребностей-чувств

Первую группу составляют «специальные» потребности-чувства, которые определяются назначением интеллекта. Они отражают критерий оптимальности деятельности управляемого объекта. Я не буду на них останавливаться, поскольку возможно бесконечное множество назначений интеллекта. Вторую группу составляют потребности-чувства, необходимые для любого более или менее развитого интеллекта и являющиеся неременным условием его деятельности. Условно назову их «рабочими». И. П. Павлов называл их «сложными рефлексам» и выделял исследовательский рефлекс, рефлекс «цели» и «свободы». Я бы добавил к ним потребность деятельности и противоположную ей — потребность отдыха. Третью группу я обозначу как «интегральные чувства». Это «Приятно» и «Неприятно»; они присутствуют в каждом частном чувстве, только по ним чувства сравниваются и суммируются, так как в принципе каждое чувство

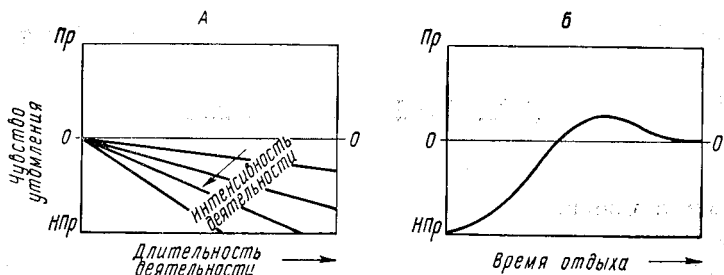


Рис. 25. Характеристика чувства утомления:

А — накопление утомления в зависимости от длительности и интенсивности деятельности; Б — динамика чувства во время отдыха.

совершенно своеобразно. В рассмотренных характеристиках чувств уже выделены компоненты *Пр* и *НПр* (см. рис. 22, 23, 24). Общее качество «рабочих чувств» в том, что они «работают» при любой деятельности интеллекта, направленной на удовлетворение любой специальной потребности и достижение специальной цели. Без них его деятельность была бы невозможна или во всяком случае весьма затруднительна. В то же время при целенаправленной тренировке эти потребности-чувства могут стать самодовлеющими и служить источником удовольствия (пример — любознательность ученого).

Несомненно, есть один общий процесс для всех видов деятельности. Это нарастание своеобразного неприятного чувства, которое выступает в роли тормоза для продолжения действий. Его можно трактовать как утомление, когда оно касается мышц, можно применить определения «скука», «надоело», когда речь идет о мышлении, но так или иначе оно присутствует. Его «входом» является сигнал от специальной следящей системы, которая учитывает длительность и напряжение деятельности, направленной к одной цели. Характеристика такого чувства показана на рис. 25. Его интенсивность является функцией продолжительности и напряжения деятельности. При прекращении действий чувство уменьшается, проходит фазы от неприятного к приятному, затем отдых надоедает, и чувство возвращается к нулю. Любое однообразное действие накапливает потребность в отдыхе или переключении на другое действие. Это касается не только всей обобщенной программы, например распознавания и исследования, но и ее частных проявлений —

опознания названия, качеств, действия, прогнозов. По каждой из этих подпрограмм ведется учет количества и напряженности одинаковых операций и соответственно нарастает потребность остановиться, переключиться или отдохнуть. Уточнение соотношения двух процессов — расходования энергии действия и накопления тормозной энергии, ведущих к одному результату — остановке действия, еще требует исследования на конкретных примерах. Торможение от утомления или однообразия — универсальный процесс, касающийся любого действия, от любых стимулов. Он необходим для всякого интеллекта.

Любознательность — вторая универсальная рабочая потребность интеллекта. Многочисленные вопросы, которые у нас возникают при виде чего-либо, являются ее проявлением. Кто-что? куда-откуда? зачем-почему? и еще масса других вопросов охватывает отношения, причинность, прошлое, будущее. Качество модели, которое возбуждает любознательность, — это ее «неизвестность», отсутствие связей с другими моделями. Когда при восприятии внешнего мира в нашей кратковременной памяти отпечатываются первичные модели объектов, рецептор отключается и включается программа распознавания. Она представляет собой поиск в памяти и сравнение одинаковых или похожих, но уже известных моделей. Извлечение моделей из памяти, их активация и сравнение — это «действие», требующее энергии, которая поступает от «центра» любознательности. Вначале нужно ответить на вопрос «Кто-что?» Затем следует установление качеств, то есть ответы на вопросы «Какой?». Качеств очень много, они выражаются «частными» моделями, заключенными в одном из кадров. Существует целый ряд структурных качеств, касающихся сравнительных размеров или формы. Есть качества, воспринимаемые отдельными рецепторами, например тепло. Динамика объекта и его отношения к другим объектам тоже могут выступать как качества.

Еще одним видом распознавания является прогнозирование. Это очень важное действие, широко используемое интеллектом. Оно сводится к действию дописывания «фразы» по известным «словам», что становится возможным лишь после опознания этих исходных слов.

75 Программа любознательности не ограничивается только распознаванием уже воспринятой модели, но

предусматривает и новое исследование среды через настройку рецепторов и нацеливание фокусированного восприятия на определенные места внешней среды или ее объект, чтобы проверить, есть ли тут важная «деталь». Исследование и распознавание продолжают до тех пор, пока иссякнет энергия чувства любопытства и стимул нейтрализуется «тормозом» со стороны утомления.

Следующая потребность как функция «рефлекса цели» — доводить дело до конечного результата даже ценой преодоления непредвиденного сопротивления объекта деятельности. Она представляется как программа слежения, воздействующая на некоторый специальный усилитель, способный дать дополнительную энергию для избранного действия тогда, когда для его окончания не хватает главного — специального стимула.

Можно это чувство трактовать иначе: как удовольствие от самого процесса деятельности, как чувство, выражающее потребность, прямо противоположную расслаблению и отдыху. Снова получается пара программ (что бывает довольно часто): потребность действовать стимулирует напряжение, расходуется какая-то энергия, возникает утомление, которое вызывает потребность в отдыхе, после чего энергия действия восстанавливается. Отдых в свою очередь надоедает, и включается потребность деятельности. Количественное соотношение обоих процессов пока не ясно. Психологический смысл «рефлекса цели» или удовольствия, получаемого от деятельности, довольно ясен. Конечно, нет уверенности, что действия ради их самих так уж необходимы для ИИ, но, может быть, этот дополнительный стимул полезен для того, чтобы доводить дело до конца.

Еще больше сомнений вызывает «рефлекс свободы» И. П. Павлова. Суть его состоит в том, что при появлении видимой помехи, препятствующей деятельности, животное (да и человек) оставляет основную деятельность и включает специальную программу уничтожения самой помехи (тогда как в соответствии с «рефлексом цели» следовало бы включить программу преодоления мешающего воздействия помехи дополнительным стимулированием главного дела). Уничтожение препятствия сопровождается «форсировкой», называемой эмоциями. «Рефлекс свободы» часто реализуется через агрессию. Ее же можно представить как эмоцию

гнева. Понимание эмоций и «экстремальных» режимов имеет значение для ИИ.

Сильно чувствуют все люди, следовательно, в принципе у всех достаточно стимулов, чтобы напряженно действовать, преодолевать сопротивление и достигать больших целей. Но в действительности это не так. Одни могут добиться желаемого, другие — нет, все зависит от характера. В модельном выражении мы определяем силу характера как способность к напряженным действиям, которая обуславливается не столько силой стимулов, сколько значимостью главного «тормоза» — утомления, скуки, «надоело». Сильные люди способны пересилить утомление, оно для них не имеет того значения, как для слабых, у которых много желаний, планов, но они не в состоянии их реализовать, поскольку быстро устают и не могут преодолеть сопротивление. Цели не достигаются, уверенность в себе исчезает, соответственно облегчаются планы и наступает детренированность механизмов системы напряжения, в частности того же «рефлекса цели». Максимум возможного напряжения — еще не все в характере, нужна, кроме того, его продолжительность: одни способны на сильное, но кратковременное напряжение, другие — на меньшее по силе, но более длительное. И уж совсем редко сочетаются оба качества силы характера: высокий уровень и большая длительность напряжения. Таким образом, характер закладывается в значимости главного «тормоза» — потребности в расслаблении, или, попросту говоря, — «чувства лени». Важную положительную роль играет ее антипод — «рефлекс цели».

Эмоции. О них следует сказать несколько слов, заканчивая раздел о чувствах. Почему-то именно они кажутся камнем преткновения на пути воссоздания разума человека. Действительно, как может «железка» переживать горе или ужас? Понятие эмоций в психологии не очень определенное. Нет четкого отграничения эмоций от чувств, даже их номенклатура расплывчатая. Все согласны, что горе, ужас, радость и гнев — эмоции, а вот любовь? Однако эмоции все-таки имеют характерные черты, хотя и не столь четкие, чтобы полностью отделить их от сильных чувств. Прежде всего они дают такую высокую активность связанным с ними моделям, что все другие модели оказываются подавленными и не могут привлечь внимание. Следствием этого является крайняя субъективность оценок:

все рассматривается через призму эмоций, человек не способен замечать другие раздражители и рассуждать здраво. Эмоции включают свою специфическую двигательную программу: гнев — потребность драться; ужас — бежать; радость — петь, танцевать; горе, — наоборот, — полную неподвижность. При этом происходит активация деятельности эндокринных органов, в частности возрастает выделение адреналина, который дополнительно возбуждает кору и играет роль положительной обратной связи. В целом эмоции производят впечатление первобытных примитивных чувств, которые включаются в условиях угрозы для жизни или по исчезновении такой угрозы. Можно думать, что на заре эволюции были простые рефлексy, потом на них наслоились эмоции, и уж потом дифференцировались чувства как более тонкие критерии поведения.

Воспроизвести эмоции в модели интеллекта не представляет труда. Нужно задать соответствующие «центры» как очень сильные усилители, «входами» на них будут чувства, «выходами» — модели раздражителя и ответного действия. Например, любая угроза вызывает потребность бежать или нападать, эмоция только многократно усиливает эту потребность вместе со значимостью раздражителя, вызвавшего угрозу. Центр эмоций включается от чувства через «реле силы» и дает дополнительное усиление соответствующим двигательным программам, тормозя при этом все остальные модели, пока действие не совершится.

Не думаю, что для искусственного интеллекта стоит задавать эмоции в таком виде. Они не улучшают, а нарушают разумную деятельность и оправданы лишь у животных, поскольку включение эндокринных регуляторов позволяет на короткое время мобилизовать дополнительную мышечную силу. Возможно, что лишь для некоторых роботов понадобится такая форсировка в условиях чрезвычайной угрозы.

Такты деятельности и напряжение

Видимо, любой интеллект, даже слабо сетевой разум человека, работает дискретно. О человеке разговор особый, а для АИ такты деятельности — обязательное условие. Для него «один такт» нужно понимать как действие по введению в кратко-

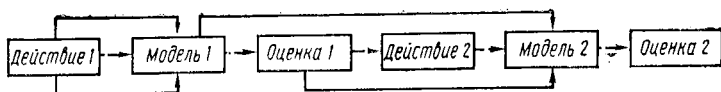


Рис. 26. Цепочка действий D с моделями M .

временную память одной новой модели или повторное активирование модели, уже бывшей в памяти. В каждый момент времени имеется одна самая возбужденная, самая активная модель. Как правило, она должна состоять из «слова» или короткой «фразы» — два-три, но не больше четырех «слов».

На рис. 26 показано, как осуществляются действия с моделями. Каждая следующая модель выбирается в результате элементарного действия, с учетом связей с предыдущими моделями, возможно, в пределах нескольких тактов. Стимул нужен для действий, но он же определяет уровень активности выбранной модели, заставляя ее генерировать энергию согласно ее характеристике «вход» — «выход», в которой учтена тренированность. На «входе» учитываются энергия стимула и других моделей (так, например, для M_2 не только от D_2 , но и от M_1).

Как отмечалось выше, действия с моделями могут иметь различный характер. Модель действий — это часть программы, и в ней тоже можно выделить иерархию: от «действий вообще» в данном направлении до конкретного действия с одной моделью (например, целая последовательность сравнений). Здесь следует ввести понятие «напряжение» как некоторый уровень активности, средний для определенного периода деятельности интеллекта. Напряжение определяется прежде всего активностью чувств или уровнем потребности, стимулом. Уже говорилось, что значимость потребностей различная: одни «жизненно необходимы» (например, страх в минуту опасности), другие (к примеру, любознательность) несут вспомогательные функции, и даже при максимальном «информационном голоде» стимул к поиску новых раздражителей нельзя сравнить со стимулом к поиску пищи или спасения.

Другой фактор напряжения — «сопротивление» объекта действия в момент его выполнения. Это понятие прежде всего касается внешней среды и подразумевает физическое сопротивление эффекторам (напри-

мер, нагрузка, которую должны преодолеть мышцы, с тем чтобы действия достигли цели). Но напряжение интеллекта необходимо и для решения трудной задачи, когда поиск в памяти нужных моделей наталкивается на большое сопротивление ведущих к этим моделям связей, если они давно не использовались. Конечно, как в том, так и в другом случае для того, чтобы напряжение возникло, нужны стимулы соответствующей силы, поскольку, если их не будет, действие просто не состоится. При этом приходится учитывать «рефлекс цели» как дополнительный стимул, который начинает временно действовать тогда, когда на пути к достижению цели, сформированной за счет умеренной потребности и неспособной вызвать большое напряжение, появляется препятствие, ставящее под угрозу ее осуществление. В таких случаях сопротивление как бы усиливает саму первичную потребность.

Время

Время является той независимой переменной, которая, как и пространство, всегда участвует в деятельности интеллекта. Значительную часть своих программ он относит к будущему, несколько меньшую — к прошлому, большую часть — к настоящему. Интеллект должен иметь чувственное ощущение времени. Настоящее метится особой «буквой», указывающей на то обстоятельство, что в данный момент времени внешние или внутренние рецепторы работают. Прошлое воспринимается по действию извлечения из памяти, а будущее — по специфическим действиям с моделями настоящего и прошлого. Течение времени фиксируется переключением кадров в моделях кратковременной памяти. В частности, мысли о будущем можно представить как операции с «фразами», в начале которых стоят «слова», отражающие настоящее, то есть с подключенными рецепторами. Затем рецепторы отключаются, и «фраза» продолжается путем «дописывания» ее начиная от «слова» — настоящего уже с «буквой» — «будущее».

Время движется только в одну сторону. Это четко отмечается в последовательности любых «слов», «фраз», отражающих события, даже во «фразах» исследования объекта типа показанного на рис. 17, где как будто движение идет по пространственным струк-

турам, но в действительности — параллельно и во времени. Время отмечается во всех действиях (последовательность вызова моделей из памяти, восприятия из внешней среды или из тела). Поскольку «буквы» действий всегда присутствуют в модели структур, то эта «метка» и создает отсчет времени, его направление. Операция обращения к прошлому или к будущему — это тоже действие, которое метит модели своей «буквой».

Временные отношения сложны для понимания (настоящее — это данный момент. Но, говоря «в этом году», тоже подразумевают настоящее). Модели, выражающие время, так же подлежат обобщению, как и любые другие. Один кадр восприятия внешнего мира — одна элементарная модель — соединяется с буквальным понятием «настоящее». Обобщение времени, то есть удлинение единицы его измерения, создание «крупноблочной» модели, производится одновременно с обобщением измеряемой этим масштабом деятельности или воспринимаемых во времени моделей. «Действие обобщения» касается времени и объектов деятельности. «Весь год пишу книги» — здесь обобщены действие и время. Нельзя обобщать их порознь, отрывать одно от другого. Как говорилось, «действие обобщения» состоит в том, что модели последовательно сравниваются друг с другом, выделяется общее из «букв» и «слов», и потом эта полученная при первых сравнениях модель сопоставляется со всеми последующими. Это касается моделей объектов и действий. Со временем проще — единицы его измерения одинаковы, поэтому обобщение времени — это использование иерархии его единиц. Операции с большими отрезками времени стали возможны только благодаря речи и счету.

Обобщение моделей действий по времени — основа долгосрочного планирования и предвидения. Поскольку интеллект может мыслить только короткими фразами, длительное предвидение обязательно связано с обобщенными моделями, которые могут развертываться в более частные единицы последующими операциями (об этом речь пойдет дальше).

Отношение событий во времени определяется по масштабу времени двух параллельно протекающих процессов, выраженных отдельными «фразами». Интеллект осуществляет слежение за многими объектами, поскольку существует несколько типов рецепторов

и следящих систем. Слежение за временем, однако, единое. Поэтому для соотношения времени разных «фраз» по их началу, концу, продолжительности нужно сравнить их маркировку по времени.

Продолжающееся и законченное действие — понятия, тоже имеющие отношение к моделям времени. Первое — это продолжение действий, начатых в прошлом и еще не отмеченных «меткой» окончания, второе — законченное, для которого она уже есть. Обычно это касается действия, состоящего из повторяющихся однородных актов (например: «Я шагаю уже час»).

Операции переключения — настоящее, прошлое, будущее, операции обобщения по времени, определения отношения начала и конца обобщенного действия к данному моменту — позволяют выразить все типы грамматических времен, используемых речью.

Реальность

Еще одно важное понятие — реальность. Реально то, что есть, что воспринимается рецепторами, следящими за внешней средой, за телом, за самим разумом. Модель с «буквой» действующего рецептора реальна. Она реальна даже в том случае, когда извлекается из памяти, если в модели образа присутствует эта «буква» рецептора. Наоборот, реальность будущего проблематична. Однако не всякое будущее можно назвать нереальным. Его реальность прежде всего определяется вероятностью будущего события. Если событие произойдет непременно, его следует воспринимать как абсолютно реальное. Другое дело, если вероятность меньше единицы: реальность события тем более сомнительна, чем меньше вероятность. Но это не все.

Реальность имеет прямое отношение к чувствам. Чувства вызываются «платой», удовлетворяющей потребность. «Плата» в настоящем очевидна: она или есть, или ее нет. Плата в прошлом тоже вполне реальна: она отмечена памятью на чувство. Будущая «плата» сильно возбуждает чувства и является основным стимулом для действий. Если бы интеллект стимулировался только в настоящем времени, то его деятельность была бы весьма ограничена. Мы все время работаем «в долг», нас заставляют напрягаться буду-

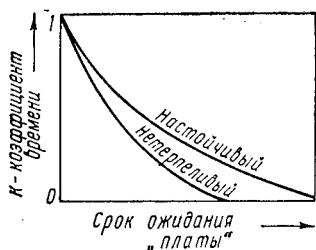


Рис. 27. Изменение коэффициента времени в зависимости от предполагаемого срока ожидания «платы» для двух типов характера.

щие удовольствия от удовлетворения потребностей, но не в одинаковой степени.

Чувства возбуждаются будущей «платой» с поправкой на реальность ее получения. Голодный человек, не имеющий никакой надежды получить пищу, не станет затрачивать силы на ее поиск, он будет просто страдать. Стимул появляется тогда, когда есть надежда удовлетворить напряженную потребность в результате направленной деятельности, то есть при наличии вероятности получить за усилия «плату».

Величина стимула от будущей «платы» определяется двумя факторами. Первый и главный — напряжение потребности и ее значимость. На рис. 22 показана характеристика «плата» — чувство. Исходное состояние потребности определяется точкой *a*, показывающей остаточную ее удовлетворенность после полученной в прошлом «платы». Приращение чувства, которое стимулирует деятельность, — это теоретическое желание $\Delta\mathcal{U}_{\text{макс}}$ от точки *a* до полного удовлетворения потребности как чувства — до предела притязаний. Абсолютная величина $\Delta\mathcal{U}_{\text{макс}}$ зависит также от значимости самой потребности.

Следует заметить, что истинное приращение чувства $\Delta\mathcal{U}$, то есть стимул, при очень напряженной потребности меньше, чем максимально возможное. Голодный человек мечтает не о максимуме наслаждений вкусными яствами, а о том, чтобы удовлетворить острый голод какой-нибудь пищей. Однако этой поправкой на уменьшение притязаний в первом приближении можно пренебречь.

Второй фактор (или коэффициент при стимуле) — это реальность получения «платы» в результате деятельности. Она определяется вероятностью успешности действий и некоторым коэффициентом времени. Представить этот фактор не просто. Можно привести мно-

жество примеров, но я ограничусь двумя. Человек переходит дорогу и видит мчащийся на него автомобиль. Откуда берутся силы! Тут коэффициент времени равен единице, так как главное не в том, что вероятность смерти высокая (но не стопроцентная, ведь человек знает, что не все умирают под колесами), а в том, что она угрожает немедленно. Другой пример. Достоверно известно, что курение угрожает человеку раком легкого. Тоже не всем, но с вероятностью 1:10. Однако не теперь, а в будущем, для молодого, может быть, лет через сорок. И юноша не бросит курить, даже если он врач и хорошо осведомлен о вреде курения. Дело снова в коэффициенте времени. На рис. 27 изображена гипотетическая функция величины этого коэффициента в зависимости от срока ожидания «платы». Показаны две кривые — одна для людей, живущих «сегодняшним днем», другая — для дальновидных и настойчивых. Коэффициент времени или поправки на будущее у них существенно разные. Видимо, этот коэффициент не является простым следствием уровня интеллекта, а заложен от природы, в генах. Интеллект только уточняет вероятность будущих событий, то есть определяет другой компонент реальности. У животных характеристика временного коэффициента падает очень круто, не говоря уже о том, что они не могут прогнозировать события далеко вперед.

Реальность будущего и оба ее компонента являются неременной принадлежностью любого интеллекта, это один из краеугольных камней всей гипотезы. Он столь же важен, как понятия о критериях-потребностях, управляющих действиями с моделями, или механизм обобщения.



Функциональный акт

Любой интеллект функционирует дискретно. Если говорить точнее, то это сочетание непрерывных и дискретных процессов. Впрочем, существуют ли вообще чисто непрерывные процессы? Во всяком случае, в сложных системах любое непрерывное есть только статистика большого числа отдельных событий. В мозге, например, вся деятельность нейронов выражается отдельными импульсами.

Дискретность внешней деятельности интеллекта я выражаю термином «функциональный акт» (ФА), понимая под ним подготовку и выполнение последовательности движений, направленных на достижение цели. Целью является новое состояние объекта управления, выраженное его моделью, которая создается в процессе самого ФА или задается извне.

Простейший ФА состоит из трех этапов: восприятие — оценка — действие. В действительности даже у животных эта цепочка длиннее и представлена, по крайней мере, пятью элементами (рис. 28), причем каждый элемент расчленяется в свою очередь на несколько простейших действий.

В самом общем виде ФА можно описать так.

Первый этап — восприятие. Рецептор, настроенный на некоторый фрагмент среды, дает его первичную модель — картину.

Второй этап — анализ. Он состоит из трех моментов:

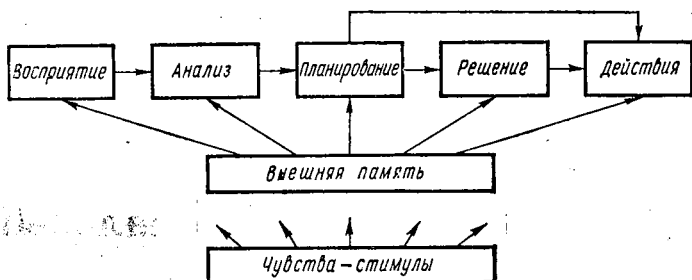


Рис. 28. Этапы функционального акта.

а) распознавание — сравнение первичной модели-картины с моделями-эталонами разной степени обобщенности, взятыми из постоянной памяти. В результате сравнения получается «вторичная модель» — переписанная своими моделями-«словами» картина среды, в которой отражена субъективность и ограниченность интеллекта;

б) прогнозирование будущих изменений среды — новая «фраза», дописанная по «словам» вторичной модели;

в) оценка — активирование чувств распознанными моделями объектов. Чувства зависят от состояния удовлетворения соответствующих потребностей к моменту начала ФА. В результате формируется суммарный стимул и выбирается «первичное действие» — модель действия в самой обобщенной форме.

Третий этап — планирование. «Первичное действие» развертывается в три «фразы» моделей, представляющих собой план:

а) «фраза» последовательности действий;

б) «фраза» эффекта — модель изменения среды в ответ на действия вплоть до достижения цели;

в) «фраза» необходимых усилий или «фраза» чувств.

Четвертый этап — решение, представляющее собой включение плана в действие. Решение возможно только в том случае, если суммарный стимул больше суммы «тормозов», предполагаемых в процессе выполнения действий и учтенных при планировании. «Тормоз» определяется сопротивлением среды.

Пятый этап — действия по реализации плана, то есть «считывание» модели последовательности действий под контролем обратных связей, воспринимаемых

изменений объекта и затрачиваемых усилий. При согласовании и недостатке стимулов действия прекращаются, и производится новое планирование.

Алгоритм упрощенного функционального акта

Реальные ФА очень сложны. Сложны модели среды, многочисленны критерии, разнообразны варианты действий. Чтобы представить алгоритм ФА, нужно упростить его до предела. На рис. 29 показана схема такого упрощенного ФА.

Рассмотрение его начнем с критериев (потребностей, чувств). Как минимум для понимания ФА необходимы четыре критерия. Первый (главный) — специфический, например голод как потребность «тела», второй — любознательность, третий — «рефлекс цели», четвертый — универсальный «тормоз» — утомление. Три последних критерия рабочие. Для каждого критерия необходимы характеристики и точки исходного состояния на них. Для голода, например, это будет зависимость чувства от количества пищи. Исходное состояние — некоторая низкая степень насыщения, оставшаяся после предыдущего приема пищи. Для критерия любознательности характеристика отражает потребность в информации, то есть зависит от числа и сложности новых моделей среды. Исходное состояние — некоторый «информационный голод». Критерий цели стимулируется от реальности цели, ее близости во времени и в соответствии с «процентом» выполнения плана. Вначале, естественно, он не действует. Критерий утомления включается от любого действия — в зависимости от его утомительности.

«Мысленные» этапы ФА — анализ, планирование — менее утомительны, но они надоедают, наскучивают. «Двигательные» этапы зависят от сопротивления объекта действия — «тормоз» от утомления может быть очень велик. По каждому виду действий — с моделями или с объектами, при восприятии или планировании — утомление отсчитывается от нуля. Значимость каждого критерия различна и задается заранее. Так, голод намного важнее, чем любознательность и удовольствие от достижения цели, но утомление в своем крайнем проявлении сравнимо с голодом.

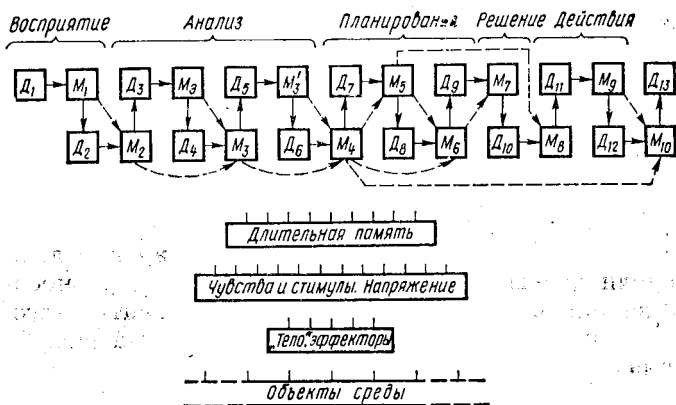


Рис. 29. Алгоритм функционального акта.

Восприятие. Восприятие среды осуществляется рецептором, для настройки которого необходимо действие. Его мы обозначим D_1 . Исходный стимул для него черпается из рабочего критерия любознательности, обеспечивающего некоторый минимальный уровень усилий. Для простоты будем считать стимул достаточным для такой настройки рецептора, чтобы получить модель среды.

В результате восприятия, как программы действия, в кратковременную память вводится модель внешней среды M_1 , например одного предмета или пространственной ситуации из нескольких объектов. Каждый из них представлен «фразой», состоящей из нескольких «слов» в кадрах памяти (см. рис. 17). Одно «слово» — обобщенная модель, другое — простейшая структура. В модель вводятся «буквы» настройки рецептора и координаты объекта.

D_2 — приведение объекта к стандартным размерам — осуществляется автоматически, за счет того же любопытства и не требует большого напряжения. Получается новая модель M_2 . Стимул при этом несколько уменьшается.

Анализ. D_3 начинает этап анализа и представляет собой программу распознавания приведенных моделей M_2 путем вызова из внешней памяти моделей-эталонов M_3 , похожих на воспринятые. Для этого используются первые «буквы» каждого «слова», как в алфавитном словаре.

D_4 — выбор наиболее похожих из вызванных моделей. Они выписываются в кратковременной памяти рядом с M_2 с последующим автоматическим сравнением и определением степени сходства. В результате получается новая «фраза» — модель распознанного объекта, переписанная собственными «словами» с указанием степени сходства, — M_3 . Любопытность (источник энергии) во время этих действий значительно уменьшается и одновременно нарастает «тормоз» — «надоело». Распознавание может быть продолжено, но для упрощения задачи ограничимся этим.

D_5 — оценка распознанного объекта как возможная «плата» для удовлетворения специальной потребности, в нашем случае — голода. Специальная «фраза» M'_3 , с которой связана модель распознанного объекта, вызывается из постоянной памяти и указывает степень удовлетворения потребности. Если ее подставить в характеристику главной потребности, отложив от исходной точки (см. рис. 22), можно получить максимальный стимул $\Delta\dot{U}_{\text{макс}}$. Он определяет основную энергию для всех последующих этапов ФА. Если насытившемуся человеку предложить невкусную пищу, то есть низкую «плату», то стимул $\Delta\dot{U}$ будет очень мал. ФА закончится на первых этапах — рассмотрение и анализ, которые отработаны на стимуле «любопытность». Человек не захочет напрягаться из-за ерунды (об этом еще будет разговор). Сейчас предположим, что стимул весьма велик.

Планирование. D_6 — определение цели. Принцип утилитарности интеллекта предусматривает автоматическое рассмотрение любой внешней картины с точки зрения ее использования для повышения удовольствия, то есть для удовлетворения потребностей. Важнейший стимул (голод) дает достаточно энергии для первой прикидки использования объекта. Суть действия состоит в создании модели цели. Зрительно это выражается в модели объекта в том виде, каким он должен стать в результате воздействий интеллекта (например, часть структуры объекта должна быть оторвана и перенесена к «телу» или попросту — взять и откусить). Такая цель может быть выражена короткой «фразой» — часть объекта, его перемещение в пространстве к новому месту. Откуда возьмется новая модель — цель? Она есть во внешней памяти в разных вариантах «фраз», связанных с моделью «съедобного» объекта. Для нахождения такой модели нужны

входные данные: обобщенная модель объекта и действия с ним. Энергия для извлечения модели из внешней памяти и переноса ее в кратковременную память черпается из стимула, определившегося в предыдущем действии. При этом стимул несколько уменьшится за счет утомления. Таким образом мы получаем новую модель — цель M_4 .

D_7 — выбор движения для достижения цели. Его модель M_5 может быть представлена на разных уровнях обобщения. Самая общая модель — это «взять», «достать» как основное движение с «буквой» обобщения. Такая модель тоже имеется во внешней памяти в виде «фразы», где связаны модель-цель и модель-движение. Движения задаются моделью рецепторов, заложенных в органах движения.

D_8 — определение сопротивления движению со стороны объекта — представляется как частная модель M_6 некоторого качества объекта. Найти ее можно по основной модели объекта или по сочетанию этой модели с моделью действия.

D_6 , D_7 и D_8 представляют собой самое упрощенное выражение этапа планирования. По существу, после каждого из этих действий должен производиться перерасчет стимулов, но для простоты эту процедуру мы объединим в одну операцию.

D_9 — новый перерасчет стимулов и «тормозов» для определения их суммы M_7 , который подготавливает важный этап — принятие решения. Предыдущие три действия были связаны с расходом энергии стимула, подсчитанного в D_5 на основании величины потребности и «платы», которую может обеспечить объект, воспринятый и распознанный на первых двух этапах. За время планирования этот стимул изменился: во-первых, за счет утомления от самой процедуры расчетов, во-вторых, за счет уточнения «ценности» объекта, которая была связана с уточнением цели, в-третьих, за счет будущих усилий для запланированных движений. Последний пункт особенно важен: мы не начинаем действий, если не уверены в том, что они нам под силу.

Принятие решения. Подсчетом стимулов заканчиваются «мысленные» этапы ФА. Решение занимает промежуточное положение: с одной стороны, действия еще не начаты, а с другой — их начало уже определено, и для этого обеспечены стимулы. Они должны быть значительно сильнее, чем для предыдущих этапов,

поскольку призваны обеспечить высокое напряжение моделей движений, способное преодолеть действительное, а не воображаемое сопротивление объекта, например мышечное усилие для поднятия тяжести. Важно понять, что во время восприятия и анализа стимул $\Delta\mathcal{C}$, как функция неудовлетворенной потребности, может быть достаточно большим, но он не используется, поскольку напряжение нужно лишь для преодоления сопротивления связей, чтобы вызывать модели из внешней памяти.

Однако такое положение бывает не всегда. Стимул для действий, в том числе и для мыслительных, то есть «чистых» действий с моделями, расходуется на преодоление сопротивления связей при введении новых моделей из памяти и их активации. Это сопровождается утомлением. Оно сильно возрастает, если процесс мышления требует большого напряжения, длится непрерывно и не достигает успеха (нет решения задачи). Удовлетворения потребности при этом нет, наоборот, она активизируется и стимул все более возрастает. Соответственно возрастанию стимула и противостоящего ему утомления увеличивается напряжение при низком УДК. Такое положение возникает при интенсивной умственной работе или в трудных жизненных ситуациях, вызывающих сильные чувства и не предлагающих легкого решения.

D_{10} — решение. Внешне это короткий акт — включение в действие ранее составленных планов, но он весьма значителен. В момент решения необходимо скачкообразно повысить напряжение, чтобы активировать модели плана M_8 , поскольку нужно преодолеть сопротивление действию со стороны объекта. У человека это связано с сокращением мышц, то есть большим расходом энергии. Решение вызывает особенно сильное напряжение, когда оно бесповоротно, когда уже невозможно остановить начатое действие (пример — прыжок через ров, разрез кожных покровов при операции).

Действия. D_{11} — действия в буквальном смысле слова: сокращение мышц у человека либо включение двигательных или манипуляционных устройств у ИИ.

Действия — это считывание модели плана. Она может быть задана в двух вариантах: в виде структурного образа объекта внешней среды как модель-цель или в виде последовательности ощущений с рецепторов органов движения. В первом случае обратной

связью является орган зрения, воспринимающий объект и сверяющий изменения, которые он претерпел, с моделью цели действий. Во втором случае рецепторы с мышц сами дают обратную связь. Однако как в том, так и в другом варианте само считывание состоит в активации моделей, непосредственно управляющих органами движения. Сами по себе они не представлены в рецепторных зонах, а выражаются только через рецепторы органов движения, но тем не менее это элементы интеллекта, такие же, как и другие. Сложные двигательные акты состоят из последовательности автоматических элементарных движений («слова» и «фразы» из «букв»).

Активность моделей действий должна быть настолько велика, чтобы органы движения, управляемые от них, были способны преодолеть сопротивление объекта воздействия. От модели плана, записанной в кратковременной памяти, поэтапно включаются модели элементарных движений, управляющие эффекторами. Одновременно рецепторами «глаза» и «мышц» воспринимается эффект производимых движений. Как и всякое восприятие, он тоже записывается в кратковременную память в виде модели M_9 . Таким образом, в кратковременной памяти активно функционирует несколько параллельных моделей: «старые» модели плана, «новые» модели объекта, измененного действием, и модели с рецепторов «мышц», показывающие их сокращения и усилия. Не представлены, но есть модели самих движений — как последовательность активации эффекторов. Эти модели, доведенные до высокой активности многократными повторениями, в дальнейшем могут стать основой автоматических сложных движений. Такими являются любые хорошо заученные двигательные акты, например произнесение слов. Разговаривая, мы просто «включаем» модели слов, а орган слуха воспринимает произносимое и выполняет роль обратной связи (в отличие от имитации чужого произношения, когда мы «считываем» звуковой образ).

Модели выполненных действий, по крайней мере две из них — изменение объекта и мышечных усилий, дают материал для чувственного контроля действий, корректируют планы получения «платы» и утомления.

D_{12} — подсчет уровня чувств, стимулов и «тормозов» после выполнения первого этапа движений

(предположим, что весь план был разбит на 2—3 этапа). Для этого нужно определить, в какой степени модель измененного действием объекта M_{10} соответствует модели — цели данного этапа действий. Недовыполнение плана уменьшает реальность всего ФА, снижает активность потребности и уменьшает стимул. С другой стороны, уже определено, каково действительное сопротивление и насколько оно соответствует «силам» эффектора. От этого зависит действительное утомление, то есть величина «тормоза», которая вычитается из стимула для определения их суммы. Если план выполнен, этапный эффект получен, а сопротивление оказалось меньше предполагавшегося, стимул возрастает и следующий этап может быть выполнен быстрее. Если же это невозможно вследствие особенностей объекта, ФА в целом может вызвать в дальнейшем большее удовлетворение. При более сильном сопротивлении «тормоз» может полностью нейтрализовать стимул и возникнет необходимость в прекращении действия, ФА останется незаконченным. Однако в этом случае включается другой «рабочий» стимул — «рефлекс цели». Невыполнение плана, отдаление цели или появление обстоятельств, угрожающих ее осуществлению, являются для него «платой» (довольно странной, но это так), уменьшающей чувство приятного, но побуждающей к деятельности, как и всякая угроза. Он суммируется с положительным стимулом надежды на истинную «плату», позволяет пересилить «тормоз» и продолжать действия, то есть перейти к их следующему этапу, который осуществляется по тем же принципам.

Уровень душевного комфорта в процессе выполнения действий зависит от их эффективности. При затруднениях он оказывается ниже предполагавшегося, если же сопротивление было переоценено, — то выше.

D_{13} — завершение ФА. Оно сводится к подведению итогов, определению окончательных чувств и УДК, но в этом и состоит его важность.

Предположим, что план выполнен полностью, «плата» получена, и это прежде всего резко изменяет главное чувство. Точка на характеристике потребности (в нашем примере — голод) перемещается в зону «приятного» и соответственно стимул для действия уменьшается до нуля. С другой стороны, накопившееся утомление представляет неприятный компонент чувственной сферы и понижает сумму чувств, то есть

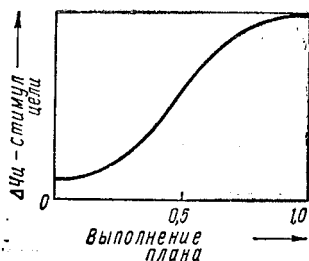


Рис. 30. Характеристика «рефлекса цели»:

$\Delta U_{ц}$ — стимул цели в зависимости от степени выполнения плана.

УДК. Третий, снова приятный, компонент дает «рефлекс цели». Он тем значительней, чем больше было преодоленных трудностей. Характеристика этого компонента показана на рис. 30. В ходе выполнения ФА «утомились», или «прискучили», или «насытились» обе рабочие потребности — любознательности и цели, следовательно, их значимость уменьшилась. Так закончился ФА, представленный в самом упрощенном виде.

Воспроизведение простого ФА в алгоритмическом интеллекте, мне кажется, не будет трудным. Потребности и их чувства, то есть критерии, задаются в виде «центров» со своими характеристиками и постоянно находятся в оперативной памяти. То же самое касается центров — моделей настройки рецепторов. Они всегда обладают хотя бы минимальной активностью. Направление рецепторов на цель и их дополнительное активирование включаются как действие, модель которого («куда направить взгляд») всегда имеется в оперативной памяти, поскольку она часто используется. Компоненты этого действия — настройка и активация рецепторов. Действие дает «первичную картину», которая тут же перекодируется цифровым кодом по особой подпрограмме. Так получается цифровая первичная модель. Этап «анализ» сводится к извлечению цифровых моделей из длительной памяти и сравнению их с первичной моделью. В результате создается вторичная модель. Планирование осуществляется по тем же принципам. Пересчет чувств, УДК, определение стимулов и «тормозов» производится после каждого этапа. Построив планы и получив достаточный суммарный стимул, АИ «принимает решение» — включает считывание плана действий. Для этого цифровая модель перекодируется в сигналы, управляющие органами движения — эффекторами. Они изменя-

ют структуру объекта согласно «модели цели», полученной при планировании. Контроль за изменениями структуры объекта осуществляется рецептором зрения, напряжение мышц при работе оценивается специальными рецепторами, полученные картины кодируются и сравниваются с планом. В промежутках между отдельными движениями пересчитываются чувства и стимулы. «Отработанные» модели находятся в оперативной памяти до тех пор, пока активность их снижается, согласно характеристике, до определенного порога, после чего они стираются. Результат ФА в виде основных моделей — «первичной картины», последовательности действий и картины объекта после осуществления ФА — переносится в длительную память. Основная трудность алгоритмизации даже простого ФА состоит в перекодировании пространственной структуры объекта.

«Круги» восприятия

В предыдущем описании ФА был предложен наиболее легкий процесс распознавания моделей — по полному совпадению с эталонами. В действительности это не так. Сложные структуры, воспринятые рецепторами и запечатленные в кратковременной памяти как первичная модель, почти никогда не имеют во внешней памяти точного аналога. Поэтому распознавание всегда носит вероятностный характер. Модель объекта представлена в нескольких кадрах, в которых запечатлены его образы с разной степенью обобщенности и некоторое количество деталей. Сравнение идет по нисходящей — от самых обобщенных моделей до подробностей. При этом требуются усилия — стимулы, поскольку нужно вызывать из памяти все новые модели — эталоны. Усилия утомляют, накапливаются «тормозы», стимулы любознательности ограничены, поэтому сравнение идет только до уравнивания стимулов и «тормозов». При этом не достигается полное распознавание объекта со всеми его качествами, а происходит лишь приблизительное опознание. Как уже говорилось, есть механизм (программа) учета степени вероятности. Этот показатель представляет «вход» для любознательности, так как всякое нераспознанное — в целом или в частностях — является новым и интригующим. Во внешней памяти у

ектов с разным набором деталей, имеющих разное значение для разных целей. Поэтому возможно запоминание нескольких вариантов «своих моделей», соответствующих «первичной модели» с разной степенью вероятности.

Какая точность нужна? И все ли необходимые для ФА детали содержатся в первичной модели?

Точность — это детали и качества. Некоторые из них являются значимыми качествами-критериями. Именно они воздействуют на центр потребности, определяют величину возможной «платы», которую представляет объект. Пример: зеленое или спелое яблоко — качество степени съедобности, в разной мере значимое для сытого и голодного человека. Его нужно точно распознать. Отсюда следует, что предел необходимой точности исследования и распознавания определяет задачу поиска качеств, значимых для данной потребности. Для удовлетворения любознательности достаточны новизна и точность сами по себе (скажем, рябина, а не калина). Для специальных потребностей нужны «избранные» детали. При первом нецеленаправленном осмотре их можно и не заметить.

Итак, первый «круг» восприятия и распознавания действует за счет стимула любознательности и дает нам приблизительную «первичную модель», которая выявляет вероятностное совпадение с не очень подробными моделями-эталоном. Второй пункт анализа как этапа ФА направлен на определение ценности объекта для удовлетворения специальной потребности и выдвигает новые задачи. Степень точности модели может оказаться недостаточной, потому что специальная потребность (чувство) предусматривает некоторые значимые для нее детали. Как только ориентировочно, по неточным моделям, определяется ценность объекта, сразу же начинается второй «круг» — целенаправленное исследование его. Возбуждение специфической потребности («голод») при восприятии объекта вызывает добавочное действие после D_4 . Оно состоит в том, что из внешней памяти по связям (адресам) потребности вызываются детали, которыми должен обладать объект, чтобы удовлетворить «голод». Во временную память вводятся «фразы» вариантов объекта, в разной степени ценных как «плата». С ними сравнивается имеющаяся уже первичная модель и обнаруживается, во-первых, ее недостаточ-

ность, а во-вторых, выявляются нужные детали, которые следует поискать в объекте, чтобы повысить вероятность «платы». Для этого включается новое действие — настройка рецептора с целью специального поиска нужной детали. В результате исследования появляется «усовершенствованная» первичная модель, которая снова подвергается анализу. Распознавание определяет, есть ли в ней искомая деталь, и если таковой нет, это понижает «ценность» объекта как «платы», что сказывается на последующем перерасчете чувств и стимулов. Модель цели на следующем этапе тоже может измениться, поскольку изменилась первичная модель.

Прогнозирование и определение динамики

Эти термины относятся к анализу, мы их опустили для простоты в описании ФА, предположив, что объект и среда совершенно статичны. В действительности же интеллект имеет дело с динамикой. Факт динамичности устанавливается глазом человека при первом восприятии, когда за доли секунды он определяет «первую производную» изменений положения или структуры объекта. Для алгоритмического интеллекта, возможно, понадобятся две первичные модели вместо одной и автоматический механизм их сравнения для определения динамики.

Однако и неподвижные в данный момент объекты необязательно неизменны в продолжение тех отрезков времени, в которых проектируется ФА. Поэтому динамика, прогнозирование изменения объекта, является неперменным действием анализа. Ее необходимость выявляется при распознавании: уже тогда, когда первичная модель заменяется, «переписывается» вторичной, в ней заложена возможность и вероятность динамики. Динамичность маркируется особой «буквой».

Прогнозирование состоит в том, что модель объекта представляется «словом» «фразы», в которой записана вероятная последовательность изменения структуры объекта или его пространственного положения в среде. Соответствующие «фразы» хранятся во внешней памяти и вызываются в кратковременную память специальным действием. Я не буду вводить дополнительные обозначения, поскольку они недопустимо ус-

ложняют схему. Распознавание будущего по «фразе» динамики носит вероятностный характер и состоит в определении промежуточных и конечного состояний объектов — в виде его новой модели, которую и нужно оценить с позиции значимости как «платы». Это не так просто.

Динамика и прогнозирование сталкивают нас с новым фактором, новой переменной — временем. При описании простейшего ФА мы не учитывали время. На самом же деле, оценивая динамику объекта, интеллект определяет время, в течение которого объект сохраняет значимость в качестве «платы», удовлетворяющей потребность. Пример: зеленая ягода. Поскольку сейчас она в пищу не годится, ФА не состоится. Другой пример: автомашина идет по дороге, которую нужно перейти. Она представляет угрозу, оценка уровня которой прямо определяется динамикой и расстоянием до автомашины. «Плата» обычно ожидается только в будущем, но она уже сейчас определяет действия. Как уже говорилось, любой развитый интеллект всегда работает на будущее. Но ценность объекта в будущем и сейчас не одинакова, и, следовательно, не одинаково его значение как стимула действий.

Для расчета динамических систем введено понятие «реальность будущего». Это очень важное понятие для любого интеллекта. Мы его разлагаем на два компонента: вероятность будущего события и «коэффициент будущего». Уже на этапе анализа реальность вводится в расчет для определения величины «платы», а следовательно, и стимула. В D_5 нужно ввести поправку для динамических систем: умножить стимул на «коэффициент реальности», полученный как произведение вероятности будущего события и значимости будущего — «коэффициента времени», рассчитанного по кривой, показанной на рис. 27. Стимул может оказаться существенно меньше или даже исчезнуть совсем.

В этапе «планирование» время присутствует постоянно. Модель цели учитывает динамику объекта («где и когда его можно взять?», «в каком он будет состоянии?» и обязательно — «сколько для этого потребуется времени?»).

D_7 — модель движений для достижения цели — полностью зависит от времени. Здесь выступает тесная связь времени и сложности действий. Простые

движения быстры, сложная последовательность их напротив, продолжительна. Не будем пока говорить о том, как планируются сложные движения, ограничимся констатацией того факта, что они требуют определенного времени для своего выполнения. Следовательно, они отдалают получение «платы», уменьшают «коэффициент будущего», увеличивают усилия и связанное с ним утомление. Все это сказывается на D_9 — пересчете стимулов после планирования.

Планирование всегда является прогнозированием, следовательно, оно вероятно. Этот фактор присутствует в планах. Степень вероятности достижения цели в результате действий оценивается по подробности исходной модели объекта и по наличию в памяти соответствующих моделей, полученных из предыдущего опыта. Эти модели-«фразы» извлекаются из постоянной памяти в то время, когда рассчитывается последовательность движений и их результат — как «фразы» плана. Результат характеризуется определенной вероятностью.

При этом может оказаться, что исходная модель объекта снова была недостаточно полной. Одно дело — модель распознавания и модель для оценки и другое дело — модель для воздействия на объект. «Фразы» действий могут потребовать новых деталей, а для этого понадобится новый, третий, «круг» исследования — снова настройка рецептора на искомую деталь, снова ее восприятие, новая первичная модель и анализ — уже с позиций выполнения действий.

Итак, при D_9 — пересчете стимулов — учитывается поправка на реальность достижения цели как функцию вероятности и требуемого времени. Нужно помнить, что само планирование несет с собой утомление, как всякое однообразное действие; это ограничивает его продолжительность и заставляет скорее переходить к решению, не добиваясь стопроцентной вероятности результата при расчете получения «платы».

Иерархия функциональных актов

Юноша пять лет учился в институте — это ФА продолжительностью в пять лет, который он планировал на такое время, исходя из существующих правил обучения. У него была цель — получить

высшее образование для работы, призванной в свою очередь удовлетворить набор потребностей, существующих к моменту поступления. Этот длинный ФА будет разделен на пять более коротких — по годам учения, те — на еще более короткие — семестры и т. д., вплоть до уроков и заданий на завтра. Такая совокупность ФА разной продолжительности и разных частных целей при одной общей — окончании института — представляет иерархию ФА. Вся жизнь человека состоит из подобных сложных ФА, направленных на удовлетворение различных потребностей и рассчитанных на неодинаковую продолжительность. Подчиненные ФА выступают при этом в двух образах — как самостоятельные и как компоненты главного.

Так мы подошли к применению принципа обобщенности к ФА. Обобщение модели относилось к моделированию внешнего мира: обобщенная структура из крупных блоков, обобщение повторяющихся объектов во времени или по этапам исследования, «буква обобщенности» как маркировка этого вида действия с моделями, даже если оно происходит автоматически в результате процесса забывания незначачих деталей.

Действия, как сокращения мышц и как операции с моделями в смысле вызова их из постоянной памяти, сравнения, прогнозирования, представлены в интеллекте моделями. Следовательно, они тоже могут обобщаться.

Обобщенные модели можно себе представить как «неясные», или «крупноблочные», без деталей, но имеющие «букву обобщенности». Она указывает, что модель получена путем действия с множеством моделей, а не через восприятие рецепторами с плохой настройкой. «Буква обобщенности» такая же, как «буква настройки рецептора», обе они показывают, как получена соответствующая модель. Обобщенная модель всегда известна, то есть имеет связи с другими моделями. Если нам нужно вообразить «собаку вообще», мы представляем какую-нибудь одну из множества виденных собак, обычно не очень ясно. Обобщенный образ собаки проявляется, когда мы пытаемся ее нарисовать, а не просто вспомнить. Он же проступает скрыто, неявно при распознавании конкретной собаки. Явная обобщенность выступает в интеллекте только через речь. Об этом еще будет разговор.

По такому же принципу возможно обобщение модели действия. Например, действие «взять» состоит из набора нескольких конкретных движений, изменяющихся в зависимости от объекта деятельности, но имеющих нечто общее — фиксацию предмета и перенесение его к субъекту. В конкретном выражении это проявляется двояко: или в работе некоторых мышц, принимающих участие во всех реализациях действия «взять», или в зрительном образе объектов, фиксированных рукой и перенесенных ближе к себе. Обобщенное действие «учиться» представить труднее, потому что оно составлено из очень разнообразных конкретных действий. Их круг расширяется по мере развития ребенка. Обобщение происходит после появления ряда моделей, имеющих общие «буквы».

Обобщенный ФА становится возможным только при развитии памяти и программы обобщения. Его условием является способность запоминать длительные периоды времени с возможностью повторения моделей и последовательного сравнения их с целью обобщения. Так получаются «фразы», охватывающие периоды времени. Как говорилось, структуры «фраз» таковы, что сами по себе они не длинны, а удлинение периодов времени достигается обобщением и созданием иерархии «фраз», расшифровывающих обобщенную модель новой короткой «фразой».

Обобщение изменений среды во времени, тем не менее, не исключает ее оценки как «платы», направленной на удовлетворение потребности. В результате оценки длительного будущего появляется стимул к действию. Обобщенной модели изменений среды в пространстве или во времени соответствует обобщенная модель действия. Эти модели «среда — оценка — действие» хранятся во внешней памяти в вариантах с разной степенью обобщенности.

Как возникает обобщенный ФА? Восприятие конкретного объекта сопровождается анализом его первичной модели, в который входит и прогнозирование. «Фраза» прогноза может охватывать большой отрезок времени в случае, когда объект изменяется медленно. Если изменения не только длительны, но и многообразны, то их обобщенная модель (из двух — четырех слов) разворачивается в целый набор коротких «фраз», описывающих каждое обобщенное «слово». Так прогнозируются этапы изменений внешней среды, ее «расписание» во времени, пространстве, качествах.

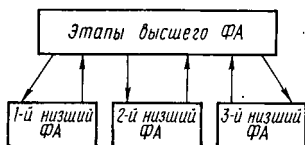
Воздействия или качества прогнозируемого изменяющегося объекта оцениваются с «коэффициентом реальности», формируются конечная цель и обобщенный стимул. В памяти хранятся модели обобщенного действия и их развертывания в частные действия применительно к каждому этапу изменений объекта. Они оцениваются по «трудности», которая выступает как «тормоз» тоже в обобщенном варианте с «коэффициентом реальности». Производится суммирование стимулов и «тормозов», и если сумма оказывается положительной, принимается решение начать обобщенный ФА. Истинное начало выражается во включении первого подчиненного ФА, то есть действий, направленных на достижение первой цели, точнее, промежуточной цели.

Проектирование такого подчиненного ФА осуществляется по общим принципам — или после принятия решения по обобщенному, «главному» ФА, или до этого, когда при предварительном планировании обобщенного ФА подробно расписывается его первый этап.

Выполнение иерархии ФА имеет свои особенности. Они в значительной мере определяются степенями зависимостей между ФА разных уровней. Например, проведение хирургической операции как цепи этапных ФА целиком определяется конечным результатом всей операции. Наоборот, в труде рабочего цели годичного или месячного периодов могут быть малозначущими, а определяющим уровнем является рабочий день или отдельные трудовые циклы с частными целями.

Для того чтобы выполнялось любое действие, необходимы стимулы. Для простейшего ФА мы их уже определили. Основные стимулы — в цели, в удовлетворении специального критерия, а рабочие критерии (чувства, любознательность, цель, утомление) играют вспомогательную роль. О поправке на реальность тоже уже говорилось. Подчиненный ФА при очень большой зависимости от главного функционирует преимущественно за счет его стимулов. Конечно, при этом действуют рабочие стимулы — «рефлекс цели» для достижения промежуточного рубежа, утомление от действий, требующее временного отдыха, и любознательность, поскольку в ходе деятельности производятся исследования. Таким образом, постоянно действующий стимул ФА верхнего уровня периодически

Рис. 31. Схема двухуровневого функционального акта. Включение очередного низшего ФА происходит после выполнения предыдущего.



сильно уменьшается за счет утомления, и действие может остановиться на некоторое время после очередного этапа, если условия среды это позволяют. В результате отдыха действие главного стимула восстанавливается, и включается следующий этапный ФА.

После каждого такого этапа производится перерасчет не только стимулов собственно этого ФА, но и главного. Учитывается возросшая вероятность достижения последней цели, увеличивается значение «коэффициента будущего» и «рефлекса цели», поскольку цель приближается во времени.

Схема двухуровневого функционального акта показана на рис. 31.

Условия успешного окончания длительных ФА особенно трудны. Во-первых, их планирование менее достоверно, поскольку ниже вероятность достижения цели из-за трудностей долгосрочного прогнозирования среды как в смысле ее «платежеспособности», так и «сопротивления». Во-вторых, нелегко оценить свои силы на продолжительное время вперед. В модельном представлении это выражается характеристиками потребностей — как «основной», так и «рабочими», например утомляемостью. Вследствие такой неопределенности длительного ФА возможно прекращение деятельности на любом его этапе, когда баланс стимулов и «тормозов» сместится в сторону последних. Правда, часто остановка имеет временный характер, если главным «тормозом» является утомление. После некоторой паузы оно уменьшается, и стимулы вновь побуждают действовать. Но бывают «тормозы» иного характера, к ним мы и перейдем теперь.

Сеть функциональных актов

Развитый интеллект тем и отличается от простого, что он управляет сложными системами по многим или, по крайней мере, несколькими критериям. Это касается только «специальных» потреб-

ностей, поскольку «рабочие» одинаковы для всех интеллектов. Например, у животных «специальные» потребности представлены инстинктами: самосохранения, продолжения рода, стадным. Каждый из них объединяет несколько составляющих: инстинкт самосохранения — потребности в безопасности и пище, стадный — потребности в общении, в самовыражении, стремление к лидерству, альтруизм. Так, даже у животных набирается с десятков потребностей, у людей же их гораздо больше. Для удовлетворения каждой потребности не обязательно нужны отдельные ФА (к примеру, общение и самовыражение объединяются), но для большинства других ставятся разные цели. Со всем нередко потребности прямо противоречат друг другу, например, такие, как безопасность и голод. Конкуренция между ФА за время, силы и средства имеет место всегда. Это означает, что ФА, направленный на удовлетворение одной потребности, должен быть тормозом для другого ФА, ведущего к иной цели. При этом компромиссы далеко не всегда возможны, поскольку условия деятельности не позволяют беспрепятственно разделять действия во времени и пространстве. Такое разделение возможно только при суммировании коротких ФА за большие промежутки времени, да и то не всегда.

Множество потребностей создает дополнительные трудности в планировании и выполнении ФА. С одной стороны, сама идея ФА предусматривает доведение его до конца, до завершения, а с другой — деятельность, направленная к одной цели, ведет к недопустимому перенапряжению другой потребности, длительное время не получающей удовлетворения. Следовательно, интеллект должен учитывать все потребности, когда планирует ФА для удовлетворения одной из них. Потребности, не получающие удовлетворения, как правило, являются «специальными» «тормозами» в противоположность «универсальному» «тормозу» — утомлению или скуке. Такое положение требует значительного усложнения расчетов суммы стимулов и «тормозов» на всех этапах ФА.

Уже при первичной оценке внешней среды — на этапе анализа — выступают противоречия потребностей. Среда удовлетворяет их неравномерно, поэтому в каждый момент одни потребности напряжены, другие пассивны, третьи активно насыщены. Оценка производится главным образом по той потребности, кото-

рая наиболее остра; другие потребности выступают в качестве «оппонентов», ограничителей. Если внешние раздражители активируют несколько неудовлетворенных потребностей, реализации которых служат разные ФА, то альтернативы возникают уже при выборе самого обобщенного «первичного» действия. Сравнение производится по суммарному приращению интегральных чувств $\Delta (Pr - HPr)$, поскольку трудно сравнивать чувства сами по себе вследствие их специфичности. Часто оценки потребностей оказываются равноценными на этапе анализа, тогда выбор самого выгодного ФА откладывается до планирования. Планы конкурирующих ФА приходится прорабатывать до той степени подробности, пока не определятся преимущества одного над другим за счет выявления значимых качеств либо за счет преимуществ большей реальности цели или меньшей утомительности действий.

Оценка среды и планирование «действий» на данный момент затрудняются не только наличием нескольких неудовлетворенных потребностей, но еще и планами ФА высших иерархических уровней (пример: студенту хочется пойти в кино, и это возможно, если учесть, что завтра преподаватель не будет спрашивать, но «нужно учиться». И он садится за книги). Высший ФА не только дает стимулы для низших, удовлетворяющих данную потребность, но он же — источник «тормозов» для тех ФА, которые направлены на реализацию других потребностей, даже если эти последние — «сиюминутные» и не имеют конкурирующих ФА высшего этажа. Следовательно, планы высших ФА нужно включать в расчет стимулов и «тормозов». Эти планы находятся в памяти и связаны с расписанием внешней среды. Слово «расписание» подчеркивает закономерность независимых от интеллекта изменений среды во времени таким образом, что в одни периоды действуют раздражители, представляющие «плату» для одних потребностей, в другие периоды — другие раздражители, направленные на другие потребности.

Деятельность интеллекта, обладающего многими потребностями, в разнообразной среде представляет собой сеть из ФА. Они подразделяются по направленности на разные потребности и по сложности иерархии. Есть среди них длительные ФА, есть минутные. Они перемежаются таким образом, что между отдель-

ными этапами длительных ФА, которые обусловлены «расписанием» среды и накоплением утомления, включаются короткие ФА, направленные на другие потребности. В результате получается очень сложная динамика действий-поступков, которую и можно назвать поведением.

Но дело не ограничивается лишь переключением действий. Это только видимая сторона деятельности интеллекта. Есть еще и другая.

«Мысли»

Большинство ФА не доводится до действий, а останавливается на этапах анализа и планирования. Эти этапы — «мысли». Таким условным названием можно обозначить последовательное возбуждение новых моделей при введении их в кратковременную память или повторную активацию их при следующем действии. «Мысли» быстры и относительно легки, так как эти модели требуют небольшой энергии стимулов, достаточной для того, чтобы только ввести их в кратковременную память (по сравнению с активностью, необходимой для возбуждения мышц, преодолевающих сопротивление среды).

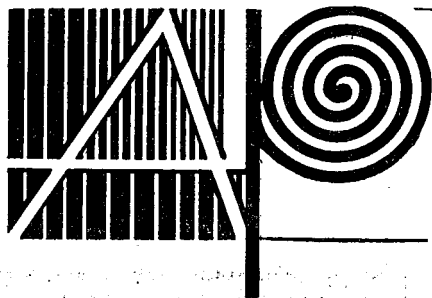
Какую же модель считать «мыслью»? В кратковременной памяти одновременно находится много моделей, даже очень много, с разной степенью активности. Видимо, «мыслью» должна быть одна из них, самая активная. Такой будет каждая новая или вновь активированная модель, и только она одна. Но почему новая активнее уже введенных? Это совсем не легко представить при количественном, модельном подходе к делу. Например, были нужны «сильные» модели для сокращения мышц, которые после выполнения своей задачи отключаются. По законам динамических характеристик активность их снижается постепенно, и эти «отработанные» модели должны бы быть более активными, чем новые модели, поступающие во время отдыха, без всякого напряжения. Все это трудно согласовать без введения дополнительных ограничений. Мысль в каждый момент только одна — это мы знаем, наблюдая себя. Впрочем, это не довод.

Рассмотрим алгоритм формирования кратковременной памяти в связи с ФА. Как уже говорилось, модели в ней образуются от восприятия среды или вводятся из внешней памяти через действия по акти-

только выбран один, другие уже не нужны. Значимость моделей в кратковременной памяти должна все время проверяться по их связям с потребностями. Связи надо прерывать, как только этапы ФА уходят вперед. К сожалению, выдвинуть подобные пожелания легче, чем выполнить.

Поиск и «выбраковывание» из памяти ненужных моделей приводит нас к необходимости специального уменьшения их активности, то есть введения торможения. Оно должно дополнить динамические характеристики моделей (см. рис. 18), отражающие постепенность самостоятельного затухания активности возбужденных моделей. Торможение сделает их более крутыми. Оно всегда имеет место в сетевых моделях и в мозге. Другим выходом является создание специального алгоритма изменения характеристик для тех моделей, которые уже использованы. Пока трудно сказать, что выгоднее для АИ.

Второй вопрос — о доминировании «главной» модели над всеми прочими. Уже было сказано, что «новая» модель должна быть «мыслью» и для этого она должна быть активнее всех остальных. То же касается и «старой» модели, если она снова понадобится и активизируется повторно, будучи вовлеченной в действия. В разделе об иерархии и сети из ФА было указано на множественность одновременно происходящих или готовящихся действий, призванных удовлетворить различные потребности интеллекта в изменяющейся внешней среде. Как в этом хаосе обеспечить целенаправленные действия, чтобы раз начатый ФА доходил до конца, а не останавливался посередине, будучи прерванным конкурирующим ФА? Конечно, есть «рефлекс цели», но он работает на все ФА и не может стать достаточно сильным механизмом, обеспечивающим приоритет главному ФА. Суть этого приоритета состоит в том, что модели, вовлеченные в главный ФА, должны активно направлять его вперед, то есть вмешиваться в выбор адреса модели для следующего переключения и даже в выбор самого действия. Нужно примирить противоположные требования: обеспечить «консерватизм», то есть доведение до конца начатого, и «гибкость» как способность переключиться на новое, если этого настоятельно требует внешняя среда или изменившиеся потребности. В свое время — в 1963 г. — я предложил для этой цели специальную программу.



Сознание и подсознание

Система усиления-торможения — СУТ

Принцип СУТ для сетевого интеллекта состоит в том, что каждый такт деятельности разума начинается с пересчета активности всех моделей, затем активности сравниваются и наиболее активная модель получает дополнительное усиление, а все другие — торможение. Обычно дополнительно активируется та модель в сети, которая наиболее значима, поскольку именно она главным образом получает энергию от моделей чувств-потребностей. Все другие, наоборот, снижают свою активность. Гипотезой было предусмотрено, что после краткого усиления «главной» модели связь к ней от СУТ как бы «устает», и усиление отключается. Одновременно отключается и торможение всех других моделей. Начинается новый цикл: снова пересчитываются активности моделей, снова выбирается наиболее активная, к ней приключается СУТ, усиливает ее и тормозит все другие. Таким образом устанавливается приоритет для самой значимой модели, которая в то же время имеет возможность усилить к следующему такту связанную с ней другую модель и, следовательно, обеспечить движение активности по моделям, представляющим этапы ФА. Повторное включение одной и той же модели исключается ее блокировкой на несколько тактов. Эта гипотеза была неоднократно воспроизведена в наших моделях сетевого интеллекта на ЦВМ и в модели,

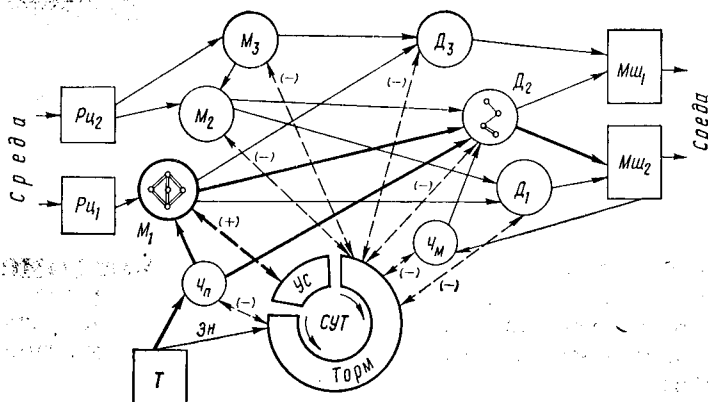


Рис. 33. Схема системы усиления-торможения — СУТ:

Pc_1 , Pc_2 — рецепторы; M_1 , M_2 , M_3 — модели образов внешней среды; $Чп$ — чувства-потребности; D_1 , D_2 , D_3 — модели действий; $Чм$ — мышечное чувство; $Mш_1$, $Mш_2$ — мышцы, T — тело; $Эн$ — источник активности для СУТ; $Ус$ — усиливающая часть СУТ; $Торм$ — ее тормозящая часть. Жирная пунктирная линия — усиление избранной для сознания модели; тонкие пунктирные линии — торможение остальных моделей; жирные сплошные линии — особенно большая активность между моделями, предполагающая захват СУТ следующей моделью D_2 . Для M_1 и D_2 показаны условные схемы нейронных ансамблей, составляющих модели.

выполненной на физических элементах. Схема, объясняющая принцип СУТ, показана на рис. 33.

Гипотеза о сетевом разуме с СУТ позволяет дать модельную трактовку психологических понятий. Вот как они выглядят:

1. Мышление — взаимодействие моделей, направляемое чувствами и СУТ.
2. Мысль — модель, усиленная СУТ в данный момент.
3. Сознание — движение активности по значимым моделям, усиленным СУТ, отражающим важнейшие отношения в системе субъект—среда.
4. Подсознание — взаимодействие моделей, ослабленных СУТ. Оно обеспечивает подготовку моделей для сознания, распознавание заученных образов и выполнение привычных движений.

Таким образом, впервые в модельном исполнении продемонстрированы сферы сознания и подсознания. Взаимодействие обеих сфер обеспечивает противоположные и взаимно дополняющие свойства интеллекта человека — дискретность сознания и непрерывность подсознания.

В зависимости от суммарного напряжения чувств изменяется уровень усиления и торможения со стороны СУТ и соответственно изменяется соотношение значений сознания и подсознания. Не следует думать, что гипотеза дает приоритет подсознанию, а СУТ только регистрирует то, что достигнуто бессознательно. В действительности после пребывания в сознании модель получает мощный толчок активности, и хотя она не может сразу вернуться в сознание, поскольку связь ее с СУТ блокирована на несколько тактов, но однако передает свою энергию другим моделям, связанным с нею, и таким образом как бы направляет дальнейшее движение мысли. Поэтому наблюдается «связность мышления». Если бы сознание являлось лишь орудием подсознания, то переключения СУТ были бы беспорядочными и целенаправленная деятельность стала бы невозможной. Сетевой интеллект с СУТ позволяет воспроизвести многие феномены мышления и поведения человека, но использование этого принципа для моделирования достаточно сложного интеллекта оказалось нереалистичным. Поэтому вернемся к алгоритмическому интеллекту и посмотрим, насколько в нем применима и полезна СУТ.

СУТ в алгоритмическом интеллекте

В принципе АИ может обойтись и без СУТ, если установить одну линию действий, движимых одной потребностью (см. рис. 32), с использованием простых «одноэтажных» ФА при отсутствии других конкурирующих действий. При этом если стимулов недостаточно, ФА обрывается на мыслительных этапах, действия не происходят, и начинается поиск нового ФА. Снова восприятие, анализ; при обнаружении значимых объектов — планирование, расчеты и т. д. Интеллект может только думать и наблюдать за средой до тех пор, пока обостряются потребности и включится ФА поиска, то есть движений в среде, цель которых — обнаружение объекта, способного удовлетворить данные потребности.

Другое дело — при нескольких потребностях и наличии конкурирующих ФА. В этом случае при некотором исчерпании стимулов для одного ФА интеллект не может оставаться в покое и ждать, а должен свое-

временно переключиться на другую линию деятельности, обслуживающую другую потребность. Осуществить это можно, только пересчитывая активность моделей, составляющих содержание другого, параллельного ФА, в процессе выполнения данного. Так возникает необходимость в подсознании, то есть параллельных операциях с моделями под управляющим воздействием различных потребностей, дающих для них стимулы. Отказаться от этого — значит лишить интеллект гибкости, способности реагировать на изменение обстановки.

Параллельные действия с моделями можно осуществлять в едином алгоритме, используя простое переключение с одной программы на другую. Переключение управляется определенными данными, например той же активностью моделей или потребностей. Создание таких программ одинаково необходимо при наличии или отсутствии СУТ. Только одно важное ограничение обязательно: недопустимы два одновременных действия с эффекторами. Коль скоро одно начато, оно должно продолжаться до своего завершения, или до возникновения серьезных препятствий, или до появления очень сильной конкуренции другого ФА. Для этого можно использовать «рефлекс цели», ограничив его применение только одним ФА. В этом случае мыслительные этапы нескольких ФА могут протекать одновременно, но на действие с эффекторами выходит только один ФА, и если решение в нем принято, прочие ФА не развиваются дальше этапа планирования.

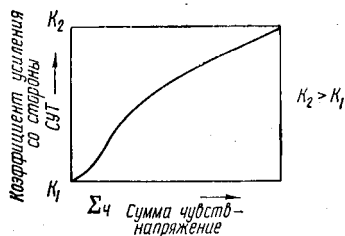
Второе ограничение касается настройки рецепторов. Как было сказано выше, существует второй «круг» восприятия, направленный на определенную цель, на поиск нужных деталей в объекте. Естественно, что если в подсознании одновременно прорабатывается несколько ФА, то может появиться необходимость в дополнительных исследованиях. Они должны быть ограничены обслуживанием одного, главного ФА. Определить его трудно, если все конкурирующие ФА еще находятся на мыслительных этапах. Нужен дополнительный алгоритм выбора главного ФА. После этого придется поддерживать именно этот ФА, поскольку для него уже проделана работа по исследованию среды. Снова неизбежно усложнение. Так постепенно приходим к необходимости все время, на любых этапах параллельно идущих нескольких ФА выбирать

один главный и давать ему приоритет перед другими с помощью специального алгоритма. Это та же СУТ, приспособленная к алгоритмическому интеллекту.

В наших сетевых интеллектах СУТ также была представлена алгоритмически: сравнивалась активность моделей, выбиралась наиболее активная и еще более усиливалась. Подобное возможно и в АИ. В нем есть кратковременная память, в которой модели имеют параметр активности. Величина активности рассчитывается по характеристике, по количеству попадающей на нее энергии от других моделей и чувств. Если это первичная модель объекта, то активность зависит от энергии, полученной при восприятии объекта, и от настройки рецептора. Так можно получить исходные уровни активности всех моделей в кратковременной памяти к началу такта, в нашем случае — к моменту действия D (см. рис. 29 и 32). Сравнивая активности моделей, выбираем наиболее активную, которая подлежит дополнительному усилению со стороны СУТ. Она и будет следующей «мыслью». Усиление избранной модели — это дополнительная энергия от СУТ, которая суммируется с энергией, поступающей от других моделей, например от чувств. Активность модели возрастает по сравнению с исходным уровнем, имевшим место к началу такта. После того как с наступлением следующего такта СУТ отключается от этой модели, падение активности будет проходить уже в соответствии с ее динамической характеристикой. Иначе говоря, энергия, полученная моделью от СУТ в тот период, когда она была «мыслью», продолжает давать себя знать в продолжение еще нескольких тактов. Это особенно важно для выбора других моделей. Согласно схеме, показанной на рис. 32, модель M_1 , побыв в сознании и получив повышенную активность к началу следующего такта, отдаст часть ее модели M_2 и действию D_2 . Это важно для них, так как повысит их шансы стать «мыслью».

Труднее воспроизвести торможение всех остальных моделей. По самой идее СУТ ее тормозящее воздействие можно представить в двух вариантах: или изменением (уменьшением) угла наклона характеристики, или прямым вычитанием некоторой доли активности, бывшей к началу такта. Выбор наилучшего варианта возможен только в конкретных моделях интеллекта. Думаю, что сперва следует пытаться уменьшать активность на некоторую долю ее исходной величины,

Рис. 34. Характеристика СУТ: коэффициент усиления избранной модели со стороны СУТ является функцией «напряжения чувств» (суммы значений чувств $\Sigma \psi$).



подсчитанной к началу такта. Тормозящий эффект влияет и на последующие такты, поскольку уменьшает шансы модели завоевать сознание и снижает влияние на другие модели. Впрочем, все зависит от степени торможения.

Усиление и торможение со стороны СУТ определяются ее характеристикой. Пример представлен на рис. 34. Степень усиления и торможения должна изменяться в зависимости от «напряжения чувств» — их некоей суммарной активности, выраженной в условных единицах. Подробности можно установить только на конкретном примере.

СУТ позволяет обеспечить приоритет начатому ФА и переключать действие на конкурирующий ФА только в том случае, когда его значимость гораздо выше первого вследствие связей с потребностями-чувствами.

СУТ и сознание

Обратимся снова к сознанию и подсознанию, чтобы попытаться выяснить их содержание в АИ. Определить сознание можно примерно так: сознание — это ориентировка интеллекта в пространстве, времени, отношениях, своем состоянии и состоянии окружающего мира. Понятно, что эти слова ничего не объясняют без их алгоритмического обеспечения. При этом возникает множество вопросов.

Какой может быть величина модели в сознании? Проверяя на себе, обнаруживаешь, что в различных случаях объемы модели неодинаковы. Когда человек рассматривает внешний мир, то содержанием его сознания является картина — ее размер, четкость, детальность. Человеческий глаз дает массу информации, поэтому видимая картина довольно сложна. Ее детальность зависит от напряжения рассматривания

и от сфокусированности взгляда. Это может быть широкая «крупноблочная» картина, охватывающая пространственную ситуацию, то есть расположение предметов, видимых нечетко. Если же зрение сфокусировано, оно дает четкую картину деталей в узком пространстве на фоне весьма расплывчатого окружения. Это примерно соответствует одному из «кадров» на рис. 17. Итак, тогда, когда СУТ приключена к рецепторному полю, в сознании — один кадр с «буквой» настройки рецептора. Последовательность рассматривания дает серию кадров со своими «буквами». В последующем эти картины переходят в память, а к рецептору приключаются новые кадры. В памяти картины быстро бледнеют, утрачивают детали и совсем стираются, если не усиливаются повторным рассматриванием или воспоминанием. Повторное восприятие запечатлевает в памяти новый кадр, но при этом устанавливаются связи между ним и предыдущим кадром, и они могут взаимно усиливать друг друга. Однако в каждый момент в сознании может быть только один кадр.

Другие рецепторы у человека дают модели гораздо более скудные в смысле разнообразия структур. Из слышанного набора звуков в сознании умещаются только короткие фразы. Можно последовательно, строку за строкой, припомнить стих, но нельзя воспроизвести сразу весь.

Таким образом, «фразы» в сознании короткие, состоят из двух — четырех «слов». Однако «фразы», бывшие в сознании в течение предыдущего такта, имеют повышенную активность, поскольку они получили «заряд энергии» от СУТ, и хотя их активность в данный такт несколько понижается торможением от СУТ, но не настолько, чтобы уничтожить эффект предшествовавшего усиления. Не исключается возможность, что некоторые «слова» могут попадать в сознание два раза подряд — вначале в конце первой «фразы», затем в начале второй. Несомненно, что между последовательностью «фраз» или «слов», побывавших в сознании, устанавливаются более прочные связи, чем между малоактивными моделями в подсознании. Тут действует принцип, в соответствии с которым связи особенно сильны между двумя моделями с высокой активностью. Так формируется «текст» — последовательность моделей, бывших в сознании и касающихся одного предмета.

При осуществлении основного ФА по программе, описанной выше, все используемые в нем модели проходят через сознание. Если бы не было параллельных ФА, а также их иерархии, не было бы нужды в СУТ и выделении понятия сознания.

Возможно, что в постоянную память переходят только те модели, которые побывали в сознании. Однако этот вопрос еще подлежит исследованию.

Функции подсознания

Значение подсознания в АИ должно быть весьма велико. Прежде всего в подсознании хранятся активные модели, уже побывавшие в сознании и представляющие значение для интеллекта, для его дальнейшей деятельности. Эти модели нужно запомнить, в последующем они являются материалом для вспоминания и повторного исследования. К ним относятся прежде всего «планы» различных ФА высших уровней. Они периодически вспоминаются и возбуждают «тормозы» или стимулы для новых ФА, обычно низшего порядка. То же самое относится к планам «соседних» ФА, приостановленных из-за «расписания» среды, в котором удобное время для осуществления следующего этапа представится в будущем. Активными остаются модели уже выполненных этапов ФА. Все перечисленные виды моделей периодически снова поступают в сознание, как только наступает пауза в действиях вследствие истощения стимулов, насыщения потребностей или обстоятельств среды, приостанавливающих деятельность. Активность этих моделей поддерживается именно повторным включением СУТ, поступлением в сознание. Другим источником энергии для них могут быть чувства, если у моделей образовались к ним сильные связи.

Повторные вспоминания картин внешнего мира могут сопровождаться новым анализом и даже планированием, чаще всего с «буквой» «нереально». Такие мысли постоянно встречаются у человека и должны быть у любого интеллекта.

Вторая функция подсознания — слежение за средой, телом, даже самым мышлением в интервалах между сознательными актами исследования этих объектов во время этапов ФА. Главное качество развитого интеллекта состоит в том, что он всегда готов от-

кликнуться на значительные изменения среды или тела, чем бы он ни был занят. Это — слежение для обеспечения собственной безопасности. Кроме того, слежение является частью любого ФА, потому что оно дает материал для обратных связей, регулирующих выполнение ФА.

Объем подсознательного слежения по сравнению с сознательным исследованием может значительно отличаться в различных интеллектах. Это определяется настройкой и активностью рецепторов, «приключенных» к объектам слежения, которые вместе определяют активность и детальность моделей, воспринимаемых и переносимых в кратковременную память. Объем слежения также изменяется в зависимости от степени концентрации внимания. При высокой концентрации СУТ очень активна (она значительно активизирует избранные модели и соответственно тормозит все остальные), и это ведет к уменьшению объема слежения. В таком случае говорят о сужении внимания.

Слежение не ограничивается простым восприятием и запоминанием. В таком виде оно не может принести пользы. Для того чтобы поднять тревогу, нужно не только воспринять сигнал, но и распознать его, и оценить меру его значимости. Таким образом, в подсознании должны осуществляться начальные этапы ФА — восприятие и анализ, результаты которых привлекут сознание, если они значительны, или останутся незамеченными либо забудутся, если не имеют значения.

Распознавание и анализ в сетевом интеллекте производятся на тех же элементах, что и запоминание, автоматически. В АИ это осуществить сложнее: необходимо включить действия с моделями, описанные при изложении содержания ФА. Нужно извлекать из постоянной памяти модели-эталоны, сравнивать их с первичной моделью, снова вызывать из памяти модели для оценки. Все это можно осуществить двумя путями: или, пользуясь большим быстродействием ЦВМ, делать паузы в обработке материалов сознания, или вводить параллельные программы для обработки информации, поступающей от каждой следящей системы. Так или иначе, это возможно, хотя и намного усложняет дело. Но интеллект не может быть простым.

117 Слежение в чистом виде не предусматривает действий. Оно дает первичный материал для сознания.

Если обнаруживается что-то важное, то это выражается в появлении очень активной модели, захватывающей СУТ, иными словами, заслуживающий внимания объект попадает в сознание. Активность моделей, полученных в подсознании, черпается из тех же чувств, которые дают активность для всякого ФА. Обычно после включения сознания информация «перепроверяется» заново — сильно активируются рецепторы, образуются точные модели, они анализируются с привлечением более точных моделей из памяти, и только затем следуют этапы планирования, решения и действия.

Однако подсознание может осуществлять и ФА с действиями, если они не мешают «главному» ФА, к которому привлечено сознание. Как правило, это очень короткие и простые ФА, в которых участвуют хорошо отработанные модели действий, не требующие большой энергии. Примеры из нашей повседневности — ходьба, почесывание, даже напевание навязчивого мотива. В таких ФА действует укороченная программа: восприятие (от слежения), распознавание, оценка (стимул), действие. Круг «автоматических» ФА ограничен как в отношении моделей распознавания, так и действий. Не исключена возможность, что модели для них постоянно находятся в кратковременной памяти, их дополнительное возбуждение невелико и легко достижимо. АИ вполне допускает воспроизведение коротких ФА, несущих вспомогательную функцию. В конце концов любое произвольное действие на низших уровнях реализуется через автоматические элементарные движения, не требующие затраты энергии и привлечения сознания. Мы автоматически ходим: повторяющиеся шаги только включаются в начале движения и останавливаются в конце выбранной дистанции. При этом следящая система в подсознании осуществляет мелкую коррекцию шагов при незначительных нарушениях рельефа пути.

Наиболее постоянными моделями в подсознании являются чувства, измеряющие меру удовлетворения потребности. Эта мера определяет уровень их активности. Следящие системы существуют для каждой потребности — они дают «входы» на модель. Это не только слежение за элементарными потребностями тела, как голод у живых существ или, к примеру, заряд питающих аккумуляторов у искусственного интеллекта. Это и слежение за мерой новизны воспринимае-

мого мира, определяющей любознательность, это и слежение за повторением однообразных действий, накапливающих скуку. Модели чувств — генераторы энергии. Если напряжение потребности велико (голод), то чувство активируется и пробивается в сознание, периодически становясь «мыслью», однако оставаться в сознании постоянно оно не может в силу особенностей СУТ. Тем не менее каждое привлечение внимания (СУТ) дополнительно активирует чувство и увеличивает потребность. В частности, это случается всякий раз, когда в процессе ФА производится оценка, по распознанной модели объекта определяется адрес потребности и к ней привлекается внимание, то есть приключается СУТ. Если интеллект ничем не занят, например действия отключены из-за «расписания» среды, наиболее острая потребность часто привлекает СУТ, чем еще больше активируется. В модельном выражении это изменение ее характеристики и соответствующее возрастание стимула. Каждый человек замечал на себе подобное: когда занят — не чувствуешь голода, а если думать не о чем, то особенно хочется есть.

Повторное привлечение внимания и усиление от СУТ могут «натренировать» даже незначущую модель настолько, что она становится очень активной и пробивается в сознание, не будучи связанной с какими-нибудь потребностями. Это явление тоже всем известно: навязчивые мотивы, картинки воспоминаний. Такая тренируемость моделей в естественном разуме заложена в их характеристиках, и, видимо, ее нужно воспроизвести в ИИ, поскольку на этом основано важнейшее качество интеллекта — самоорганизация.

Слежение за временем в буквальном смысле (за секундами, минутами и т. д.) возможно только при овладении речью. В более простом варианте время присутствует всегда: оно заложено в конкретных моделях — динамических картинах. Однако при запоминании последовательности явлений (картин) в памяти происходит «сжатие» времени за счет действия «обобщения по времени», о котором говорилось выше. Тем не менее направление времени всегда сохраняется. Видимо, это объясняется односторонней направленностью связей между моделями, то есть пространственной структурной организацией памяти, в которой воспроизводится не только статическая структура моделей, но и их динамика.

«Координаты» сознания

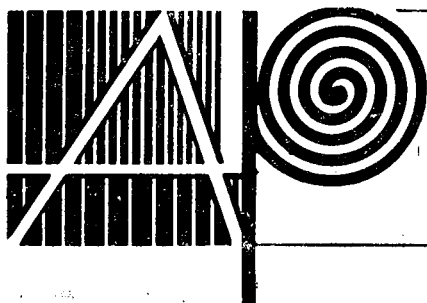
Модельное выражение словесного определения сознания, как постоянной ориентировки в пространстве, времени, отношениях, самом себе, выражается в комплексе постоянно обновляющихся активных моделей, полученных от соответствующих следящих систем. Все эти модели связаны друг с другом, так что любое действие, любой ФА не может происходить без участия основных «координат» сознания, как я называю перечисленные параметры. (В развитом интеллекте есть еще несколько других, о них речь далее). Воспроизвести координаты сознания в алгоритме не просто, но возможно. Видимо, для этого каждое действие с моделями, предусмотренное в ФА, должно пройти по всем «координатам», которые могут наложить ограничения, стать «тормозами». Такое невозможно без связей моделей-«координат» с «тормозом». Может быть, следует сделать его универсальным «нельзя», связанным с интегральными чувствами *Пр* и *НПр*. В этом случае каждое действие с моделями и их выбор направляются и ограничиваются различными комплексами моделей. В состав этих комплексов входят «координаты». Алгоритм проверки по ним необходимо воспроизвести, хотя возможно, что не нужно проверять по всем моделям, а достаточно выбирать только значимые (вопрос требует отдельной проработки).

Алгоритм интеллекта должен воспроизводить ту же динамику сознания, которую можно наблюдать на себе. Возможно, что для специализированных искусственных интеллектов это и не лучший алгоритм, но для универсального он необходим. Наши мысли в самом упрощенном варианте состоят из этапов ФА, перемежаемых воспоминаниями, отвлечений на следящие системы — на чувства, на внешний мир — в те моменты, когда ФА осуществляется автоматически или возникает перерыв в действиях, связанный с «расписанием» внешней среды. Все модели, побывшие в сознании, остаются в кратковременной памяти, так же как и связи между ними. С течением времени память подвергается «эрозии», незначимые модели, с низкой активностью, не возвращаются в сознание (не вспоминаются) и забываются. Значимые получают подкрепление в подсознании от чувств и захватывают СУТ в интервалах между этапами ФА.

В подсознании происходит колебание активности моделей, энергия переходит по связям от одних к другим, порой возникают связанные процессы — незаконченные или даже полные ФА. Видимо, следует поделить память, как кратковременную, так и постоянную, на отдельные зоны, представляющие собой «рецепторные поля», — внешний мир, тело, действия, чувства. Процессы взаимодействия моделей в подсознании в первую очередь ограничиваются пределами зон, однако периодически распространяются за них.

Отразить сложную динамику сознания и подсознания алгоритмом — трудная задача. Можно только предполагать, что она выполнима, но нельзя это утверждать, пока не будут созданы хотя бы простые модели интеллекта. А ведь то, что мы рассмотрели, несравнимо с разумом современного человека. В лучшем случае оно соответствует уровню обезьяны или несколько больше — нашего первобытного предка. У него уже была иерархия ФА, которой, видимо, нет у животных. Их планирование происходит автоматически, охватывает ограниченные отрезки времени и диапазон действий. Я намеренно не вводил понятия социального сигнала, не выделял его из воздействий внешней среды, из ее «картины». Это, наверное, допустимо у животных, на уровне первой сигнальной системы.

Можно было бы этим и ограничиться. Создание искусственного интеллекта алгоритмического типа, соответствующего уровню высших животных, уже представляет интересную задачу, поскольку при наличии обширной внешней памяти такой интеллект мог бы решать много прикладных задач управления, встречающихся в технике и организации. Но моя задача шире. Чтобы доказать универсальность алгоритма интеллекта, нужно довести его, по крайней мере, до уровня современного человека и желательно наметить пути дальнейшего совершенствования. Разумеется, надежд на воплощение алгоритма в программах будет еще меньше, но кто знает? Может быть, найдутся энтузиасты, которые захотят вложить труд в решение этой задачи. Мне кажется, что она уже по силам современной вычислительной технике.



Интеллект уровня человека

Начнем с условий, в которых действует интеллект человека. Общая его схема показана на рис. 21. Интеллект замыкается на сложную внешнюю среду, состоящую из природы, техники, общества и набора статических моделей — книг, чертежей и т. п. Главное здесь — общество как новая сложная система, предъявляющая определенные требования к индивиду.

Одно из основных отличий системы «общество» — циркуляция информации наряду с веществом и энергией. Это — модели и передача их сигналами. Интеллект индивида должен воспринимать сигналы, выражающие модели, создавать по ним свои модели моделей, выдавать их вовне сигналами и быть способным проверить, как они превращаются в модели у других индивидов. Все это охватывается понятием «речь» со всеми ее аспектами и воплощениями.

Общество, как сложная система, имеет свой разум. Не следует пугаться этого утверждения: у общества есть органы управления и модели, по которым они функционируют. То, что часть этих моделей находится в разуме отдельных людей, не меняет дела. Алгоритмы управления (или модели) все равно принадлежат обществу как таковому, в мозг людей, причастных к управлению, они поступают вторично и за счет функционирования общества как системы. Здесь

полная аналогия с отдельным человеком: его мозг функционирует не только за счет структуры из клеток, но и благодаря специфике деятельности этих самих клеток. В них заложены возможности памяти и переработки поступающих сигналов в «выходы».

Для деятельности в сообществе нужны специфические критерии в интеллекте. Они имеются уже на уровне высших стадных животных. Это доказали исследования этологов. Однако общество, как сложная самостоятельная система, потребовало дополнить биологические критерии разума, обеспечивающие эффективность совместных действий животных, более гибкой надстройкой, способной отразить специфику различных типов общественных структур. Алгоритм интеллекта высокого уровня должен предусматривать такую возможность — изменение заложенных критериев и даже создание новых.

Следующее требование — память. Она должна быть более длительной и большего объема. Без речи такая память не может дать существенных преимуществ, но речь без нее нельзя реализовать.

Человеческое мышление по своей природе социальное. Но для того чтобы проанализировать его механизмы, следует сначала ввести упрощающую модель индивидуального мышления, то есть мышления в тех его чертах, которые, по преимуществу, общи для всякого интеллекта (в том числе животного), и подняться далее от него до интеллекта на уровне общества.

Речь

Речь зародилась в процессе усложнения сигналов, с помощью которых первобытный человек воздействовал на внешнюю среду в целях удовлетворения своих потребностей. Звуковые сигналы у животных преследуют ту же цель.

Стимулы для речи — те же, что и для любого вида управления. Однако уже у животных при стадном существовании появился особый стимул — потребность самовыражения, потребность отразить свои внутренние чувства, свое состояние внешними признаками. По существу, это управление, поскольку проявления самовыражения воздействуют на окружающих, побуждая их к определенным поступкам.

123 Прежде всего задачу выражения чувств выполняют

мимика и позы (они остались и у человека). Затем к мимике прибавились звуки и интонации их произнесения. У человека, как наиболее стадного из всех высших млекопитающих, потребность самовыражения из подсобной превратилась в самостоятельную. Она и теперь является важнейшим стимулом речи. В частности, этому служит эмоциональная окраска произносимых фраз. Человеческая речь усвершенствовалась благодаря двум факторам — большой памяти и хорошо развитому речевому аппарату. Память позволяет запоминать долгую последовательность звуков, а мышцы гортани — произносить их. Уже животные обозначают звуками не только свои чувства, но и приказы к действию и информацию о некоторых важных объектах и даже их состоянии. Просто у них мал набор сигналов и беден словарь.

Звуковая речь («говорение») — это мышечное действие, такое же, как любое другое. Произнесение фраз — функциональные акты, доведенные до действий. Законы ФА распространяются на речь. Здесь есть все его этапы. Планирование, выраженное в трех «фразах»: само движение языка и гортани, предвидимый эффект и собственные чувства. Есть обратная связь — слух и дистантные органы чувств, направленные на объект и призванные уловить, как подействуют слова. Проще всего в этом плане контролировать воздействие слов-приказов: «Возьми! Принеси! Брось!»

Понимание чужой речи также осуществляется по общим правилам ФА. Это восприятие и запоминание чужих слов — первичная модель; это анализ, то есть распознавание, прогнозирование, оценка. Слово, обозначающее предмет, можно уподобить естественному признаку последнего, который не представляет модель объекта во всей его сложности, но, будучи абсолютно характерным для него, оказывается достаточным для распознавания объекта в целом. Например, животное на расстоянии узнает врага по некоторым коротким, но очень характерным звукам. Чем же от них отличаются слова — набор последовательности звуков, обозначающих предмет?

Образное и речевое мышление. В алгоритме интеллекта приходится выделять образное и речевое мышление. Главное, несомненно, образное. Оно выражается в структурах моделей, отражающих структуры объектов и их изменения во времени. Зрительные образы — чисто структурные; звуковые образы — это

волны, которые превращаются в структуры. Развитый интеллект содержит очень большой набор структур в своих моделях. Этот набор отражает не только разнообразие внешнего мира, но и способ образования моделей интеллектom (например, для одного и того же объекта может быть много моделей с разной степенью обобщенности). В основе деятельности интеллекта лежит именно взаимодействие моделей-образов, моделей, отражающих в своем большинстве действительные структуры внешнего мира. Движения представлены тоже моделями, отражающими последовательность возбуждения (активации) мышц, иными словами, это тоже отражение естественной структуры организма.

Образное мышление, то есть взаимодействие моделей внешнего мира, выражает все сложности управления объектами со стороны интеллекта. Говоря проще, взаимодействие моделей-образов составляет основу грамматики и синтаксиса речи, которые мы привыкли относить к только ей присущим качествам. Речь лишь обозначает, маркирует основные образы. Правда, тем самым действия с моделями становятся более целесообразными и эффективными.

Обратимся к грамматике — грамматике образов, а не речи.

Предмет, имя существительное как часть речи в образе выражается материальным объектом — пространственной структурой, отграниченной от остального мира. Наиболее обобщенная модель материального тела — это и есть существительное, отвечающее на вопрос «кто-что?». Действия — это модели, отражающие движение, изменения. Несомненно, есть обобщенная модель динамики как противоположности статике, неподвижности. Это глагол.

Качества — это прилагательные; их очень много, и они выделяются из первых двух классов моделей — предметов и действий. Часть качеств определяется по виду энергии, излучаемой объектами, другие выражаются в особенностях структуры или ее изменении. Все однородные модели сравнимы по параметрам, которые им присущи и которые могут быть представлены в частных моделях. Так возникают степени сравнения по качествам, сопоставление размеров или скоростей. Есть и алгоритм счета — как последовательность движений.

Думаю, что имеется модель, соответствующая местоимению, — как обобщение предмета или живого

существа, которое выступает в повторяющихся моделях-«фразах», касающихся одного объекта.

Наречия, как абстрагированное качество отношений, тоже можно представить в модельном выражении. Труднее представить понятия, вытекающие из качеств или действий: «скорость», «высота», «цвет», не говоря уже о таких сложностях, как, например, «относительность» или «неопределенность». Формирование этих понятий стало возможным уже благодаря речи.

Есть образное выражение в чувствах и для модальных глаголов — ощущения «могу», «хочу», «должен».

Вероятность, как мера совпадения сравниваемых моделей, учитывается при всех действиях с моделями. Отсюда: «может быть», «возможно».

Вспомогательные глаголы «быть» и «иметь» имеют чувственное выражение, иными словами, их модели в разуме существуют и без речи.

Спряжение глаголов связано с понятиями времени и ощущениями действия, с отношениями того и другого. Уже говорилось, что настоящее время совпадает с моментом, когда активированы рецепторы, которые воспринимают действия. Отсюда все формы существующего в английском языке продолженного времени Continuous. Например, «I am writing» и эквивалент в русском языке «Я сейчас пишу». Обобщенное время дает нам неопределенное понятие настоящего, будущего, прошедшего. Перфектные формы означают законченное действие («Я уже сделал»). Для нас это означает, что функциональный акт завершен и это получило выражение в чувствах.

Очень интересно выражение действительного и страдательного залогов. Пусть мы произносим фразу: «Злая собака укусила кошку». Можно акцентировать любое из этих слов, и их чувственное выражение изменится. Любое слово может стать ключевым в этой фразе — и не только в словесной, но и в образной. Эти акценты мы ставим, наблюдая данную сцену. Применительно к глаголу акцент выражается в залоге: действительный залог относит глагол к собаке, страдательный — к кошке. К сожалению, бедность языка не позволяет выделить в качестве «ключевого» слова ни сказуемое, ни определение. Можно, конечно, сказать «Злая-презлая собака...», этим мы выделим при помощи ключевого слова качество злости собаки.

Или еще: «Собака укусила-таки кошку» — здесь выделено сказуемое. В образном мышлении такие акценты определены активностью главного «слова» во «фразе» последовательности действий.

Склонения имен существительных и обслуживающие их предлоги — все это имеет образное выражение. Грамматические формы языка не требуют новых образных понятий, они только регистрируют сигналами речевых слов то, что врожденно задано в образах, без чего деятельность интеллекта была бы невозможна. Жалость к укушенной кошке и негодование против собаки заложены в чувствах, которые освещают зрительный образ, а речь лишь регистрирует это более или менее выразительно.

Интересно в этом плане проследить образное выражение грамматического понятия «наклонения». В английском и русском языках выделяют три наклонения: изъявительное, повелительное и сослагательное.

Повелительное наклонение выражает требование, приказ и в устной речи всегда подтверждается интонацией. Это прямое отражение управляющих воздействий интеллекта, направляемое сильными стимулами. Изъявительное наклонение скорее является отражением потребности самовыражения, без акцента на управление, и касается реальных фактов. Сослагательное наклонение отражает отношение к реальности. Я уже останавливался на этом важнейшем понятии, выступающем как качество любого события, действия, которое существенно влияет на стимулы ФА. Реальность — это одна из важнейших «координат» сознания.

Типы предложений тоже появились из образного мышления. Они известны — утверждение, отрицание, вопрос. Утверждение — положительный результат распознавания, отрицание — наоборот, а вопрос — это неизвестное, требующее исследования. Он может относиться к любой части фразы и формируется как результат анализа образа, в процессе которого производятся распознавание массы качеств, прогнозирование, вспоминание.

Всеми этими рассуждениями о грамматике я хотел показать, что существуют врожденные формы действий с моделями-образами, заложенные в структуре мозга, и что речь не придумана человеком произвольно, она лишь зафиксировала эти врожденные

программы средствами внешнего выражения. Сами эти средства, в виде слов и их изменений, действительно, придуманы, и для них нет эквивалентов в природе. Наличие врожденных механизмов действий с моделями-образами доказывается одинаковостью ряда понятий, которые выражаются во всех языках, хотя и разными словесными формами. Такие понятия, как времена, пространственные и временные отношения, реальность, выделение ключевого слова во фразе, есть во всех языках. Только поэтому возможны переводы.

Как говорилось, главное мышление — образное, а не речевое. Образные картины гораздо богаче словесных, в них неизмеримо больше разнообразия, то есть информации. Если представить себе количество моделей-образов в коре, которые одновременно «живут», проявляют активность, то разве можно сравнить их с количеством слов? Передача образов словами кажется медленной, неповоротливой и невыразительной. Человек мыслит преимущественно образами, а внутренняя речь, которую мы зачастую отождествляем с мыслями, только комментирует образное мышление. Для этого достаточно понаблюдать за собой: как мы вспоминаем вчерашний день, как планируем действия, как чувствуем, и всюду — преимущественно образы.

Получается так, что назначение речи в мышлении служебное. Однако это ни в коей мере не умаляет ее значения. Генезис общества был связан с развитием труда и общения, включая речевую деятельность. Каждый человек обязан речи своим индивидуальным развитием.

Значение речи можно свести к тому, что она служит средством общения, обучения, хранения, передачи и распространения информации. Речь позволяет экономно выражать сложные образы, хотя и с потерей информации. В сущности, речь моделирует образы, а всякое моделирование является необходимым средством управления сложностью. Почти необъятное море моделей-образов заменяется ограниченными моделями речи и знаков, тем самым появляется возможность выявить некоторые структуры, которые тонут в сдвиге образов. Наконец, речь является единственным средством обозначения, маркировки обобщенных моделей и их передачи. Понятия «существо», «вещь», «предмет», являясь обобщенными образными моделями, немыслимы без речи.

Наилучшим ли образом выполняет речь функции моделирования образов? Думаю, что нет. В этом отношении гораздо более эффективны графики и их математическое выражение — формулы. Не случайно именно они стали всеобщим языком точных наук. Речь придумана для общения как расширение набора элементарных звуковых сигналов, которыми обмениваются животные, усовершенствования довели ее до границ науки, но не сделали ее однозначной для понимания.

Все это связано с семантикой, иначе говоря, с соответствием моделей-образов моделям-слов. Общеизвестны трудности перевода с одного языка на другой и неудачи кибернетиков в этой области. Буквальный перевод часто не отражает истинного значения текста на исходном языке. В чем причина? В множественности значений одного слова и в множественности синонимов, отражающих один образ. Слова приобретают разный смысл в разных контекстах. Иначе говоря, «действующей единицей» является не слово, а фраза или по крайней мере ее часть. Более того, даже фраза, вырванная из текста, может не только изменить свой смысл, но и утратить его.

Прежде всего трудности объясняются тем обстоятельством, что первичным является образное мышление и речь не в состоянии отразить все его богатство. Но не только этим. Есть специфика самого образного мышления, которое тоже нельзя свести к простой формуле «модель — объект». Как было показано ранее, предмет воспринимается в виде целой серии кадров-моделей разной обобщенности и детальности, объединенных друг с другом «горизонтальными» связями в некий комплекс.

Механизмы сознания и подсознания еще более усложняют проблему выражения мыслей словами. На рис. 35 показано, как формируются переключения моделей через СУТ. Видно, что хотя сама «мысль» представляет собой короткую «фразу», в действительности это только вершина айсберга. Модели — «координаты» сознания активны и, не входя непосредственно во «фразу» сознания («мысль»), а пребывая в подсознании, влияют на выбор действия и следующей модели M_2 . На осуществлении такого выбора сказываются различные факторы в подсознательной сфере, в которых суммировано прошлое и настоящее: одни воздействуют на стимулы, активируя определенные потреб-

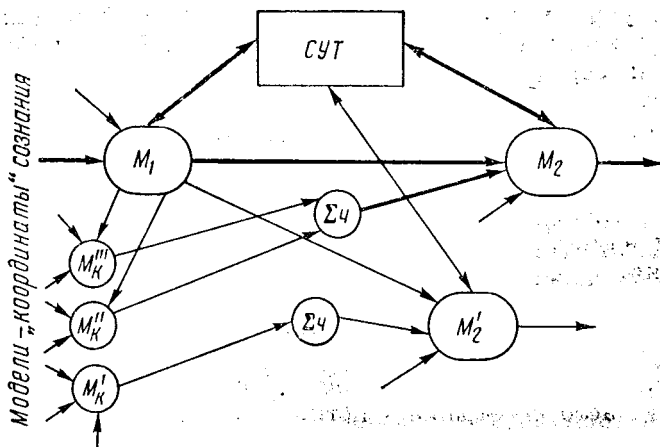


Рис. 35. Схема переключения СУТ (переключения сознания):

M_1 — последняя модель, находившаяся в сознании; M_2 — следующая модель для захвата СУТ — сознания. Ее высокий уровень активности предопределен получением активности от модели M_1 и от суммы чувств $\Sigma\mathcal{C}$, которые в свою очередь возбуждены от моделей — «координат» сознания (время, пространство, отношения и пр.) M_K' , M_K'' , M_K''' . «Конкурирующая» модель M_2' имеет более низкую активность и остается в подсознании.

ности, другие — на «тормозы», возбуждая противоположные потребности. Происходит сложный пересчет активностей, в результате которого модель M_2 получает перевес и именно ее выбирает СУТ в качестве следующей «мысли». Отсюда берет начало новое направление действий с моделями. Речь не может поспевать за такими сложными отношениями и отражает лишь следствие их, скользит по поверхности («Мысль изреченная есть ложь»).

Понимание речи. Алгоритм понимания речи сводится к первым двум этапам ФА — восприятию и анализу. Восприятие — это кодирование рецепторами слуха звуковых колебаний воздуха в последовательность сигналов рецепторных клеток, воспринимающих разные частоты. Так получается первичная модель услышанной фразы (см. рис. 14) в серии кадров, состоящих из «слов» с «буквами» интонаций и ударений. Анализ начинается с переписывания услышанной фразы своими словами и с перестановки акцентов. Это действие разворачивается как распознавание — на первый план выделяются слова, хорошо знакомые, с четкой собственной семантикой, а слова чужие, малоизвестные, либо совсем опускаются, либо заменяются

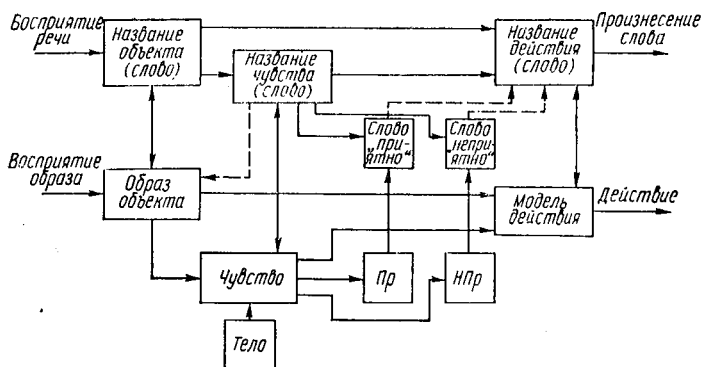


Рис. 36. Взаимоотношение моделей образов и моделей речи. Моделям образов объекта, чувств и действий соответствуют модели словесных эквивалентов. Между ними имеются прямые и обратные связи, так что они могут взаимно активировать друг друга.

сходными по значению (как кажется) своими словами. Одновременно формируется следующая строка моделей — главная; это модели-образы, вызываемые из внешней памяти по семантическому «словарю соответствия» между словами речи и образами. Здесь в полной мере проявляется субъективность: у каждого человека свои ассоциации слов и образов и своя иерархия соответствий одного слова нескольким образам. В результате образуется «фраза» из моделей-образов, как если бы она создавалась при прямом восприятии объектов среды, но, разумеется, бледнее. Параллельно с переписыванием производится оценка «слов» и целой «фразы» в виде возбуждения (активации) чувств, с которыми имеют прочные связи rozpoзнанные слова и образы.

Схема алгоритма понимания речи как отражения слов образами показана на рис. 36. Она относится к утвердительному предложению, которое просто передает информацию.

Для того чтобы слышать говорящего, не нужно много стимулов — слух работает как следящая система и автоматически передает в слуховой анализатор звуки, в том числе и речь. Однако для понимания речи необходимо не просто слушать, а слышать то, что сказано. Для этого уже нужно подключение СУТ, сознания. Она периодически переключается с рецептора на «материал речи», на модель фразы, зафиксирован-

ной в кратковременной памяти, с тем чтобы анализировать ее путем типичных действий с моделями. Именно это требует усилий и является утомительным. «Простота речи» выражается в легкости извлечения из постоянной памяти моделей-эталонов для распознавания, прогнозирования и оценки услышанного. Напротив, при восприятии незнакомого материала для понимания требуются усилия, то есть значительные стимулы, чтобы извлекать из памяти редко используемые модели, которые далеко не точно соответствуют слышимому. Все это утомляет, и слушатель теряет нить изложения, его сознание отключается от анализа слышимого и приключается к чему-то другому.

Стимулы для слушания и понимания те же, что и для любой другой деятельности: от потребностей, включая и любознательность.

Понимание речи очень субъективно и искажено, иными словами, воспринятая фраза в разной степени соответствует образам и чувствам человека, произносившего ее. Это объясняется рядом причин (см. рис. 36). Словарь синонимов, по которому чужая фраза переписывается своими словами, ограничен и неточен. На подбор слов влияет исходное состояние чувств, которые «подсказывают» для переписки слова, наиболее значимые в данный момент. Еще более субъективен словарь соответствия слов образам. Так, исходные чувства влияют на исходную структуру фразы. К сожалению, это положение трудно реализовать в алгоритме, поскольку пришлось бы постоянно переписывать «букву значимости» и «адреса чувств» для всех моделей постоянной памяти. Однако постоянные связи слов с чувствами, характеризующие данный интеллект и отличающие его от других, закладываются в постоянную память и проявляются при понимании речи. Вторая причина — субъективность оценки «фразы» образа после того, как он уже воссоздан. Это зависит от характеристик чувств и их исходного состояния.

Речь, как самовыражение, формирует новую потребность — «говорить правду». Ложь не может дать удовлетворения этой потребности — выразить свои чувства действием, то есть произнесением слов. «Чувство правды» является проявлением такой потребности. Правда — это совпадение истинной «фразы» образов, воспринятых рецепторами, и вторичной «фразы» образов, полученной при переписывании «слов

речи» образами. Совпадение («правда») приятно, несовпадение («ложь») неприятно. В связи с этими чувствами стоят слова «Да» и «Нет». Первое регистрирует совпадение двух предложенных для сравнения образных фраз, второе — их различие. В различиях может быть и количественная мера — по общим принципам вероятностного сравнения — «возможно» вместо «да» или «нет».

Доверие — это тоже одно из чувств, появившееся вместе с речью. Хотя оно и связано с чувствами симпатии и любви, которые имеют биологические корни, но его нельзя оторвать от речи. Мы доверяем информации, услышанной от авторитета, который, во-первых, знает предмет, а во-вторых, всегда говорит правду. Таким образом, реальность слышанного определяется не только материалом речи, но и источником ее.

Понимание вопросительной фразы имеет свои особенности, которые стоят в связи с действиями, вызванными речью. Как уже говорилось, услышанная речь — это такое же внешнее воздействие, как и физические факторы среды, только переданное в виде модели. Воспринятая и понятая словесная фраза вызывает определенные чувства, они могут быть стимулами к действию, а характер действия зависит от образного содержания понятой фразы. Утвердительное предложение может и не затрагивать чувств, вызывающих ответные действия, вопросительное же всегда направлено на ответ. Будет ответ или нет — зависит от силы стимулов, и не только заключенных в вопросе, но и в отношении к вопрошающему. Вопрос включает поиск в памяти, а произнесение ответа планируется и выполняется как любой двигательный акт. Он может быть заторможен и остаться только во внутренней речи.

Устная вопросительная фраза, кроме вопросительного местоимения, имеет еще интонацию, которая дополнительно выделяет ключевое слово, указывающее направление поиска и формирования ответа. Разберем самый простой пример. Скажем, меня спрашивают: «Где ты был вечером?». Любое из этих слов может быть выделено как ключевое. При переписывании вопроса своими словами оно становится на первое место и от него строится поиск. Задача простая: нужно образно представить вопрос сначала в обобщенном виде: «где» — это пространство, «вечером» — время, «ты» —

чувство себя. Производится перебор содержащихся в памяти образных моделей пространства и себя в порядке убывания степени обобщенности во времени: сначала — прошлое, затем — «вечером» (подразумевается «вчера»). В зависимости от акцентирования, от ключевого слова в вопросе, которое подкрепляется чувствами к спрашивающему, избирается направление поиска в памяти, то есть происходит активирование соответствующих моделей. Если акцентируется слово «ты», то это предполагает отделение модели себя от моделей других людей. Если ударение на слове «вечером», то усиливаются модели времени с целью их уточнения; если же выделено слово «где», осуществляется переход от обобщенной модели пространства к более детальной. Повторяю: стимулом поиска в памяти, обеспечивающим затраты энергии на это действие с моделями, являются чувства, которые вызывает спрашивающий. Так или иначе, в результате вспоминания — активации в кратковременной памяти или извлечения из внешней памяти — строятся «фразы» из моделей-образов, начиная с ответа на ключевое слово: «Я был вечером там-то», «Вечером я был там-то», «Там-то я был вечером».

Понимание отрицательных предложений базируется на тех же принципах: по словам речи выстраиваются «фразы» из моделей-образов, и отрицание воспринимается как отсутствие соответствующего образа. Здесь тоже возможно акцентирование различных членов предложения, и от этого зависит порядок «слов-моделей» во «фразе» образов.

Понимание требований, приказов, просьб, советов состоит в том, что соответствующее предложение переписывается в виде «фразы», фиксирующей требуемый результат. Задача воспринимающего приказ состоит в том, чтобы «написать» две другие недостающие «фразы» плана: «фразу» порядка действий и «фразу» усилий, которые при этом понадобятся, а также установить конечную «плату» за выполнение приказа. Поскольку в понимание речи входит и ее оценка, то производится сопоставление стимула, исходящего от приказывающего, с «тормозом», который включает в себе выполнение приказа. Одновременно прорабатывается противоположный план, и даже не один: выполнять ли вообще приказ либо выполнить его не так, как предписано. Это служит материалом для действия.

Особо следует остановиться на чтении и понимании письменной речи. Здесь нет принципиального отличия от устной речи, только канал связи другой — зрение — и сам источник сведений — книга — лишен тех качеств, которые у собеседника представляют множество стимулов для восприятия речи. Текст лишен также дополнительной информации в виде интонации и мимики говорящего, дающих те самые акценты, о которых шла речь выше. Правда, их можно заменить особым порядком слов в предложении, знаками препинания, шрифтовым выделением и, наконец, рисунками, но это всегда беднее живой речи и зрительного восприятия говорящего. На первых порах учебы ребенок читает вслух и так приучается «слышать текст». У некоторых это остается на всю жизнь, однако при беглом чтении в память вводятся только короткие звуковые эквиваленты слов. Во всем остальном понимание письменной речи не отличается от устной. Так же параллельно с чужими словами выстраиваются свои, знакомые, так же за ними следуют модели образов, вызванные из постоянной памяти по «словарю» коротких «фраз». В конце «фразы» подытоживаются чувства.

Запоминание речи. Запоминание услышанной речи происходит по общим принципам памяти на воспринятые внешние воздействия. В кратковременной памяти некоторое время сохраняются все услышанные слова. Наряду с ними сохраняются и «фразы», переписанные своими словами и образами. Именно они определяют неодинаковую значимость услышанных слов, а следовательно, их различную активность в памяти. Как правило, наиболее активны ключевые слова — подлежащие, затем, по убывающей, прочие члены предложения. То же касается и активности последовательности «фраз» в тексте — мысли можно подразделить на основные и поясняющие их. По мере забывания происходит процесс обобщения моделей услышанного, распознанного и оцененного: важное остается, неважное обобщается и заменяется соответствующими моделями. Быстрее всего забывается сама услышанная фраза — от нее остаются лишь некоторые яркие слова. Дольше сохраняется «фраза» смысла — в виде своих ключевых слов и соответствующих им моделей-образов. В конце концов в памяти остается главный смысл услышанного в виде образа и обстоятельств, при которых этот смысл был воспри-

нят, — время, место и источник речи. «Буква реальности» при таких моделях определяется доверием к говорившему.

Прочитанный текст забывается по тем же законам: сначала исчезают чужие слова, потом малозначащие свои, так что в результате остаются только ключевые слова с «буквой» обобщения. Они и заменяют всю фразу. Потом забываются малозначащие фразы, и весь текст оказывается представленным смысловыми тезисами — короткими ключевыми фразами с добавлением зрительных образов, вместе составляющими упрощенную модель прочитанного.

Как должна быть организована память интеллекта, владеющего речью и общающегося с себе подобными? Эту память можно назвать смешанной — в том смысле, что она состоит как из образов, так и из слов и фраз речи. В зависимости от содержания моделей превалирует тот или другой их вид. Слова представляют собой как бы модели образов. Во многих случаях они более экономны и являются единственным инструментом для выражения образов. Принцип различной обобщенности моделей относится одинаково к обоим их видам. Он позволяет организовать память в виде набора коротких фраз.

В словесных моделях для обозначения качества обобщенности существуют специальные слова, дающие название классу однородных моделей. В образах нет обобщенной модели женщины — «женщины вообще», а есть ряд неполных моделей конкретных женщин с «буквой», указывающей, что их много и они имеют нечто общее. Этот ряд моделей и обозначен одним словом — «женщина». Другие классы моделей в цепи обобщений обозначены специальными словами, например «живые существа», «толпа». Такая маркировка обобщенности очень расширяет действия с моделями и обеспечивает ФА высших иерархий. Впрочем, и в словесных моделях обычные слова могут выступать в обобщенном смысле. К примеру, в описании хирургической операции этапы ее обозначаются конкретными словами: «разрез», «выделение из спаек» и т. д. Под такими словами подразумевается целая последовательность конкретных движений, которую можно тоже описать словами, если «развернуть» обобщенную модель в подробностях. Для обозначений обобщения в речи иногда используется множественное число. Смешанные «фразы» (модель-образ) в соединении со

словом речи широко используются в памяти и позволяют в большем объеме и в то же время экономнее выражать информацию, чем чистые образы.

Если обратиться к отдельным объектам памяти, то можно сделать несколько предложений об их структуре.

Яркие зрительные картины запечатлеваются в виде моделей-образов. Так же выражается и их динамика. Часто серия картин перемежается словами. Так, например, запоминаются путешествия: отдельные картины живописных видов, затем словесные перечисления мест и дат. Хронология путешествия запоминается в словесной форме, но сложная последовательность образов тоже возможна. Образы можно трансформировать в слова — описать картины, и наоборот — хронологию либо названия обобщенных этапов путешествия связать с отдельными картинами или даже с последовательностями кадров, отражающих динамику. Картины в большинстве случаев неясные, но с «буквой» реальности, подтвержденной связью со словесными моделями.

Образы людей, с которыми мы так или иначе общаемся, и всего, что с ними связано, тоже запечатлены в памяти. Зрительные картины их внешнего вида, отдельных черт, динамика внешнего выражения поступков перемежается звуковыми образами речи. Однако содержание их высказываний запоминается в том виде, как было понято, то есть переписано своими словами. Хотя в процессе понимания чужой речи воспринимаемые слова обычно сопровождаются строкой моделей-образов, но эти образы смутные и не запоминаются. В памяти сохраняется словесный пересказ речи собеседника.

Запоминание научной информации, как правило, почерпнутой из книг, осуществляется в том виде, в каком она оформлена, то есть в словах, схемах и рисунках. Словесный текст обычно не запоминается и заменяется обобщенным образом. Если имеется «предметный материал» — машины, эксперименты, большие, то он помнится в зрительных картинах, смешанных с речью. Большой объем знаний ученого возможен только благодаря систематизации моделей. Систематизация — это обобщение. Знания укладываются в иерархию обобщенных моделей с «вертикальными» и «горизонтальными» связями — адресами. Параметры, по которым производится обобщение — си-

стематизация, могут быть самыми различными. Они предписаны методологией науки. Специальные слова — термины — изобретаются учеными для свертывания информации. В этом их плюсы, минусы же в том, что терминология ограничивает распространение научной информации среди непосвященных.

Произведения искусства представляются в памяти тремя элементами. Запоминаются зрительные картины или тексты, пересказанные своими словами по всем законам обобщения словесных моделей. «Буква» нереальности присутствует всегда, как и адрес источника информации.

Существует словесная память действий, такая же, как и образов. Она представляет собой словесные инструкции, как управлять эффекторами (к примеру, какие нужно сделать движения для достижения определенной цели). Есть «словарь», в котором записаны соответствия слов элементарным движениям. Он не доходит до низших уровней управления отдельными мышцами и ограничивается целостными двигательными актами: «взять то-то, так-то и сделать то-то». Движения обычно расписаны только по их эффекту. Обобщение словесных моделей движений производится так же, как и моделей образов. Оно состоит из ключевого «слова» с «буквой» обобщения или, иначе, из серии движений, выполненных различными мышцами, но направленными на одну и ту же цель: «взять», «бросить», «оставить», «остановиться». Такое действие разворачивается в серию конкретных движений в зависимости от свойств объекта и состояния мышц, о котором поступают сигналы с их рецепторов. Если, скажем, нужно взять яблоко, а правая рука занята, то мы берем его левой, свободной. Но обычно мы не произносим мысленно слова «взять левой рукой», за исключением случаев, когда предварительное разворачивание обобщенной модели не производилось.

Планы действий в ФА можно выражать как словами, так и образами. Обычно они выражаются «смешанным» языком: обобщенная модель действия — словом речи, «фразы» цели и усилий — образами. Вообще план может выражаться параллельно двумя языками: за первичными образами следует словесная модель, «комментарий» или за словесной мыслью — ее образное воплощение. Это осуществляется за счет «горизонтальных» связей между соответствующими моделями обоих кодов — образом и словом речи. Пла-

ны тех ФА, которые направлены на людей и сопровождаются произнесением слов, чаще всего выражаются словесным кодом, а те, что непосредственно направлены на предметы, — образными.

Речь и убеждения. Общеизвестно чувственное значение слов. Словесными приказами можно вызвать чувства так же, как и физическими воздействиями. Так называемые убеждения человека представляют собой не что иное, как словесные формулы, имеющие чувственное значение.

Психологи определяют их как социальные чувства, называя терминами «долг», «совесть», «честь», «патриотизм» и др. В обществе существует масса запретов, табу, правил, предписывающих, как нужно поступать в том или другом случае. Эти словесные формулы постоянно регулируют поведение человека во всех его проявлениях. В плане гипотезы о сущности интеллекта их можно рассматривать как источники энергии, такие же, как и центры естественных потребностей. Они активируют модели действий, побуждая преодолевать сопротивление среды, так же как их активирует реальная боль или голод.

Откуда черпают энергию словесные убеждения?

Мне представляется, что убеждения — это просто очень тренированные модели слов и связанных с ними образов, обладающие высокой собственной (спонтанной) активностью. Как у всяких активных моделей, у них множество связей с другими моделями, они часто используются и попадают в сознание. Такой модели свойственна большая крутизна как статической, так и динамической характеристики. Это означает, что при небольшом возбуждении на «входе» модель сильно активируется. Активность долго сохраняется, передаваясь на другие модели. В этом плане высокая активность даже биологических чувств, возможно, в значительной степени является результатом тренировки, а не проявлением врожденных, заданных характеристик. Может быть, связи в большей степени обладают качеством врожденности, чем активность нейронов?

Этот вопрос важен не только для задач создания ИИ, но и для психологии, поскольку характеристики центров чувств-потребностей определяют индивидуальные качества личности, а их изменение от внешних воздействий — воспитуемость человека. Вопрос о естественных чувствах — это сфера гипотез. То, что

убеждения целиком продукт воспитания, не вызывает сомнения. История дает сколько угодно примеров того, что целенаправленным воспитанием можно привить совершенно противоестественные нормы поведения, перечеркивающие всю биологическую сущность человека с ее инстинктами и врожденными чувствами.

На рис. 36 показана схема взаимоотношения врожденных чувств как выразителей потребностей, моделей-образов объектов внешнего мира и словесных моделей разного значения. О биологических чувствах уже говорилось — соответствующие потребности заложены в генах, хотя они изменяются воспитанием. Для простоты в схеме показано сугубо «телесное» чувство, но существуют и «информационные» врожденные потребности. Выше я определял их как необходимые для любого интеллекта. Это рабочие критерии — утомление («надоело»), «рефлекс цели», потребность в новизне («исследовательский рефлекс»), «рефлекс свободы». Входами для них является информация извне, получаемая специальной автоматической обработкой моделей-образов объектов внешнего мира. Производные стадного инстинкта, будучи биологическими, тоже имеют информационные источники возбуждения, например соревнование. На схеме для примера приведено одно частное чувство и два интегральных — *Пр* и *НПр*. Для них имеются эквиваленты — слова (голод, любознательность и др.). Активность моделей слов прямо зависит как от воздействий со стороны соответствующих чувств, так и от восприятия речи извне. В свою очередь эти «словесные чувства» дают обратную связь на «настоящие» и могут их до некоторой степени активировать, а следовательно, тренировать или подавлять. На рис. 36 показаны модель образа и модель эквивалентного ей слова. Между ними существуют двусторонние связи, причем неодинаковой проходимости. Модель-образ имеет врожденные или приобретенные связи с чувствами. Связи между образом и словами — названиями чувств — устанавливаются в процессе обучения и воспитания.

При восприятии объекта его оценка формируется не только по связи с чувством, но и по активности соответствующего данному чувству слова. При восприятии речи возбуждение модели слова передается на те же чувства, на которые поступает возбуждение при восприятии образа, эквивалентного данному слову. При

этом возбуждаются не только слова — названия чувств, но и собственно чувства. По таким связям производится оценка речи.

Как было показано в разделе о функциональных актах, модель-образ имеет связи не только с чувствами, оценивающими объект, но и с моделями действий (см. рис. 32), которые активируют образ в силу врожденных или приобретенных связей. У действий есть словесные эквиваленты, с которыми они соединены двусторонними связями, как всякие слова и образы. В свою очередь модели слов-действий связываются с моделями слов-чувств — *Пр* и *НПр* — или их аналогами «хорошо» и «плохо», «да» и «нет». «Приятно», «хорошо», «да» — это стимулирующие слова, «неприятно», «плохо», «нет» — это слова-«тормозы». Словесные эквиваленты образов связываются со словесными эквивалентами действий и чувств и формируют простые рефлексy — элементарные ФА на словесном уровне. Их структура аналогична структуре ФА на образном уровне. Но дело не ограничивается существованием двух параллельных уровней ФА — образного и словесного. Между его элементами развиваются «диагональные» связи. Это значит, что модель действия активируется не только от образа, но и от слова. ФА, включаемый от слов, не обязательно должен соответствовать образному ФА, отражающему врожденный рефлекс, стимулируемый биологическими потребностями.

Создание моделей словесных ФА достигается ценой многократного повторения слов: слово-объект — слово-действие — слово-чувство. Связи между словами в последующем действуют на уровне образов, которые им соответствуют. Так хорошо заученный «словесный» ФА превращается в «образный». Его модель, соединяющая «входы» и действия, обеспечивается прочно заученными связями.

Это и есть убеждения. Сила их — в проходимости «диагональных» связей и в тренированности словесных моделей действий, выраженной в их характеристиках. Если убеждения очень «натренированы», то они в состоянии навязывать выработанное поведение в любых условиях. Если их сила меньше, то они достаточны для того, чтобы затормозить не очень настоятельные естественные потребности.

Обычно убеждения существуют в виде очень обобщенных моделей, предписывающих, в каких случаях что нужно делать.

Формирование убеждений происходит при повторении соответствующих словесных формул, но одного заучивания недостаточно. Необходимы либо авторитет, либо доказательства истины.

Действия, обусловленные убеждениями, имеют свою специфику. Произнесение слов нужно рассматривать так же, как любой двигательный акт. Раз так, то речевая деятельность подчиняется всем законам функциональных актов: восприятие и анализ информации, стимулы планирования, решение, само действие. Особенно ярко проявляется планирование: каждый знает, что ответственные выступления мы по несколько раз репетируем про себя, прежде чем произнести вслух. То же можно наблюдать и при сложных движениях руками: при их планировании в мышцы поступают импульсы и даже осуществляются легкие пробные движения.

Первые два этапа речевого ФА — восприятие и понимание (анализ) речи — уже были рассмотрены. В результате анализа внешняя информация становится значимой и вызывает стимулы для действия. Вопрос о стимулах, побуждающих к произнесению фраз, очень не прост.

Стимулы в ФА предусматривают энергию для преодоления сопротивления объектов среды, для напряжения. Это и физическое сопротивление, и интеллектуальное напряжение, связанное с перебором большого числа трудно вспоминаемых моделей, и любое действие, вызывающее скуку своим однообразием. Произнесение слов кажется наиболее легким физически, поэтому как будто не требует значительных стимулов. На самом же деле это так и не так. Простая болтовня с приятным собеседником действительно воспринимается как отдых. Но, с другой стороны, словесный приказ,двигающий армии, вынесение приговора требуют стимулов, больших, чем, скажем, труд каменотеса. Итак, «сопротивление» речевым поступкам может быть весьма значительным и выражаться во вполне реальном ущербе для говорящего в смысле воздействия на его потребности и чувства. Именно эти последствия речи и требуют стимулов.

Любой ФА всегда направлен на удовлетворение потребности — одной или нескольких. Я перечислил их в общем виде, подразделяя на «специальные» и «рабочие». Первые являются выражением специальных критериев управления и зависят от объектов сре-



Рис. 37. Цепочка модели действия.

ды и «тела» при интеллекте, вторые должны быть неременным атрибутом любого интеллекта, если он претендует на оптимальное управление сложной системой. В качестве специальных критериев для высших животных и человека выступают биологические потребности — инстинкты. У интеллекта на уровне человека, то есть связанного речью с другими подобными существами, появляются специфические стимулы, связанные как раз с речью. Ими являются убеждения.

Стимулы — это активность модели чувства, настолько сильная, что она способна активировать другие модели, например модели действия, которые в свою очередь становятся способными активировать эффекторы вплоть до преодоления сопротивления среды. Получается цепочка (рис. 37).

Чувства-стимулы активируются в простом случае от «тела», в более сложном — от моделей среды или самого интеллекта (например, любознательность).

В случае действий, вызванных убеждениями, цепочка сокращается, поскольку выпадает стимул как критерий действий, заложенный изначально в интеллект природой или конструктором. Его заменяет просто высокая активность словесной модели действия, натренированная воспитанием.

Убеждения становятся возможными только при условии, что среда способна оказывать воспитательные воздействия и в элементах моделей заложено свойство тренируемости, то есть изменчивости характеристики в результате интенсивной деятельности. Явление тренируемости элементов, т. е. элементарных моделей, — основа свойства самоорганизации интеллекта. Самоорганизация — это изменение структуры в результате самого действия, повышающее эффективность последующих подобных действий. В случае человеческого общества — это создание убеждений как специальных стимулов деятельности, лишь косвенно связанных с биологическими качествами человека.

143 Убеждения, как большая активность моделей избранных словесных фраз, выступают в роли стимулов

только на уровне высокого обобщения — на уровне так называемых принципов. Мелкие действия по соображениям, вытекающим из принципов, — это лишь развертывание обобщенных моделей в детальные. Обобщение происходит как в отношении моделирования среды, и в частности понимания речи, так и в отношении действий по их одному определенному качеству (например, «не вредить ближнему», «не щадить врага»).

Важный вопрос — о мере активности убеждений как стимулов. Говоря иначе: сколь напряженную работу способен выполнить интеллект по убеждениям? При обычных потребностях эта мера определяется значимостью какой-то одной потребности по сравнению с другими — и напряжение, создаваемое за ее счет, уравнивается утомлением. Вот почему от голода и страха люди способны на большое напряжение, а из побуждений любознательности — по крайней мере большинство их — только на малое. Если в результате напряженного действия человек получает «плату», которая удовлетворяет потребность, то стимул сразу резко уменьшается и для последующего напряжения нужно, чтобы потребность снова обострилась.

«Платит» ли общество человеку, действующему из убеждений? В большинстве случаев — да, «платит» одобрением, удовлетворяет некоторые биологические потребности человека. Однако в принципе убеждения не требуют «внешней платы», то есть ответа среды. Напряжение деятельности, обусловленной убеждениями, определяется активностью соответствующих словесных моделей, способных преодолевать сопротивление среды, а продолжительность этой деятельности определяется универсальным «тормозом» — утомлением. Значимость убеждений в ряду биологических (или изначально заданных) потребностей определяется соотношением их максимальных активностей.

Особенности речевых функциональных актов. Главная особенность этих ФА состоит в их направленности — на другие интеллекты, владеющие речью как данной системой сигналов и моделей, надстроенной над универсальными моделями образов. Сюда относятся разговоры всех видов, т. н. «вербальное поведение». Кроме того, речь используется в коллективных действиях, направленных на материальную среду. В этом случае речь выступает в качестве вспомогательной деятельности.

Первая особенность речевых ФА определяет стимулы для речи — они находятся в сфере отношений между интеллектами. Я все время употребляю этот термин вместо слова «человеческое мышление», для того чтобы подчеркнуть, что человек — только частный случай универсального интеллекта, обладающего специфическими биологическими потребностями, которые оставила нам природа наших предков — стадных животных. Однако в любом случае, как только появится система интеллектов, взаимодействующих между собой через знаки, возникнут «общественные» проблемы, а следовательно, и специфические стимулы, регулирующие отношения между членами сообщества. Более того, можно даже наметить, в какие формы должны неизбежно вылиться отношения между членами сообщества, а отсюда — какими будут особые стимулы для этих отношений как особого вида деятельности.

Исходя из направленности речевых ФА, для их планирования и выполнения нужны модели объектов управления — других интеллектов. Самый простой вариант этого встречаем у детей. Видимо, дети (как и животные) рассматривают другого человека как объект среды, наравне с куклой, игрушкой, собакой. Действуя и наблюдая, они изучают реакции матери, близких, постепенно связывая их со словами: «сердится», «ласка» и пр. Эти реакции являются ответом объектов на действия ребенка, они удовлетворяют его потребности, вызывают соответствующие чувства, и он соответствующим образом планирует свои поступки. Естественно, что его ФА непродолжительны, так как ребенок еще не умеет создавать модели, обобщенные во времени. И у взрослых имеются импульсивные речевые ответы, такие же короткие, как рефлекторные движения. В своем полном выражении модели собеседника представляют собой перевоплощение себя в его личность («поставить себя на его место»).

В сложных речевых действиях главное — это планирование. Оно производится по общим правилам ФА — в три «фразы»: «что я скажу», «как это подействует» (вероятные чувства и реакция собеседника), мои чувства в ответ на реакцию собеседника. Эти чувства — «плата», на которую нацелено любое действие. Планирование сложного разговора предусматривает сначала создание обобщенных планов — модели собственного действия в обобщенном виде, в таком же

выражении — ответа партнера и своих чувств как реакции на этот ответ. Обобщенные словесные модели, как было сказано, представляют собой слово или короткую фразу с «буквой» обобщенности. Обобщенное планирование речевых ФА — это выбор тона разговора и ключевых слов. Когда из нескольких вариантов плана разговора и реакции собеседника выбран один, тогда начинается подробное планирование — те самые мысленные диалоги, которые мы постоянно ведем, собираясь начать серьезный разговор, затрагивающий наши чувства. При этом ключевые слова разворачиваются во фразы. Они мысленно произносятся, уже слышатся в уме ответные реплики, ощущаются чувства — в слабом варианте, планируются собственные ответы, новый круг разговора и т. д. В ходе такой проработки, как и при всяких подобных расчетах, выясняются слабые и сильные стороны выбранного обобщенного плана и иногда вносятся коррективы.

Подобные речевые ФА планируются исключительно на словесных моделях. Их материальные эквиваленты — образы предмета разговора — присутствуют где-то рядом, в самом ближайшем подсознании, и скрыто взаимодействуют со словами. Только мимика партнера и зрительные образы всей сцены будущего разговора периодически проникают в сознание, оживляя речь.

Существуют такие типы мышечных действий, которые предполагают речевое общение. Это коллективные действия, направленные на предметы. Речь в них выступает в подсобной роли, главное внимание сконцентрировано на материальных объектах воздействий. Однако планы тут сложнее, чем при индивидуальных действиях, поскольку нужно предусмотреть не только команды своим мышцам, органам действий, но и словесные распоряжения партнерам, их ответные действия и суммарный эффект деятельности. Здесь тоже необходимо перевоплощение, когда дело касается больших напряжений. Планы существенно зависят от той роли, которую предполагает выполнять субъект, — руководить или подчиняться. Личностные отношения обязательно вмешиваются в такой ФА и предусматриваются в планах.

Решение в речевом ФА — самый важный этап. Для него нужен значительный стимул, измеряемый последствиями речи, которые могут быть вполне материальными или затрагивать «чувства общения». По той же причине серьезные разговоры могут сопровождаться

значительным напряжением, высокой активностью СУТ, а следовательно, и последующим утомлением.

Само действие, как этап ФА, представляет собой считывание модели плана. В большинстве случаев заготовленные фразы не произносятся полностью. Обычно фраза импровизируется по обобщенной модели плана, представленного ключевыми словами и чувствами, которые нужно выразить. Механика речи у человека настолько отработана, что модели слов функционируют как единое целое, а не как последовательность звуков или слогов. Произнесение звуков слова — это действия, осуществляемые без участия сознания. То же касается и письма. Обучение ребенка устной и письменной речи представляет пример отработки таких моделей действий через многократное повторение. Интонации речи несут важную информацию. У человека они тоже «автоматизированы» и в наибольшей мере выражают чувства говорящего, если их специально не сдерживают.

Можно выделить несколько типов речи. Прежде всего — монолог или диалог. В первом случае партнер присутствует пассивно, его реакция, обратная связь, «сопротивление» учитываются, но без ответных речей. Во втором случае ответы оппонента (или собеседника) являются важным элементом как планов, так и самих действий.

Начнем с монолога.

Самым простым вариантом монолога является приказ. Он прямо выполняет функции управления и подкрепляется соответствующей позицией приказывающего, обладающего правом применить санкции. Они подразумеваются или даже указываются в приказе. У животных отношения отдающего и получающего приказ определяются физической силой (животные тоже отдают приказы). В человеческом обществе существует целая гамма взаимоотношений — действий и контрдействий, которые направлены на удовлетворение различных потребностей и представляют «плату» со знаками плюс или минус. Взаимодействия имеют материальную либо информационную природу, когда действуют факторы, рассчитанные на такие чувства, как любознательность, свобода и т. п. Все это представлено в моделях образов, стоящих за словами приказа.

Другой формой монолога является «рассказ» (например, пересказ событий или впечатлений о путешествии). Здесь речь представляет собой «считывание»

моделей-образов по плану, который строится в соответствии с хронологией событий или выражает некоторую идею рассказчика. Так или иначе, любой план составляется из обобщенных моделей-образов и эквивалентных им слов с «буквами обобщения». Каждому пункту плана соответствует программа его развертывания в изложении деталей. Подбор слов для выражения образов производится автоматически. Обратная связь в монологе представлена реакцией аудитории, которая воспринимается как объект воздействия. В зависимости от реакции подбираются выражения либо даже перестраивается план. При этом мысли обгоняют слова. Фраза развертывается из обобщенного образа почти механически, подсознательно, а сознание в это время приключается к плану, к ключевым словам и тону следующих фраз. Они могут меняться. По такому же типу строится лекция.

Еще одним вариантом монолога является пересказ чужой речи по памяти. В сущности его отличие от рассказа состоит в том, что здесь в основе лежат не образные модели, полученные первично от репетиторов, а свои слова и соответствующие им образы из памяти, на которые переведена чужая речь в процессе ее понимания.

Диалог — самая частая форма речи. Суть ее — обмен речевыми ФА. Диапазон диалогов очень широк — от рассказа с редкими вопросами слушателя до ссоры. Столь же велико разнообразие чувств, побуждающих к обмену фразами. ФА каждого собеседника состоит из восприятия реплики или вопроса, понимания ее (то есть перевода на свои слова), анализа (то есть распознавания образов), прогнозирования и самое главное — оценки, выяснения собственных чувств. Чувства дают стимулы для ответа и определяют его форму. Вопросы возникают в процессе переписывания реплики оппонента своими словами и служат для заполнения возникающих пробелов. Их может быть очень много — все зависит от вызванных фразой чувств и наличия моделей в памяти, позволяющих подбирать подходящие слова.

Алгоритм диалога можно представить такой схемой беседы собеседников А и Б:

А	Б	А	Б
Вопрос	Ответ.	Вопрос	Ответ.

Формирование вопросов и ответов — это ФА со всеми его атрибутами: иерархией планов, учетом реакции собеседника. В пределе при всяком диалоге предусматривается перевоплощение в личность оппонента.

Существует такое понятие — «внутренняя речь». Это слова и фразы, произнесенные мысленно. С другой стороны, мышление всегда связано со словами. Разница здесь чисто количественная: внутренняя речь более четкая, доведенная до правильных грамматических фраз, это «почти произнесенные» слова. Установлено физиологами, что при этом нервные импульсы направляются к мышцам языка и гортани, которые даже сокращаются, но лишь слегка, не издавая звуков. Фактически это заторможенные двигательные акты. В процессе же мышления отдельные слова всплывают в сознании непроизвольно, как бы комментируя мышление образами. Наблюдая за своими мыслями, можно хорошо заметить параллельность двух кодов моделей — образов и слов. Если реализуются ФА, направленные на предметы, на материальный мир, тогда в качестве главного выступает образное мышление, но слова всегда следуют за образами, особенно тогда, когда для этого достаточно времени. Наоборот, если ФА направлены на людей (на интеллекты!), осуществляется мышление словами, за которыми следуют образы.

Высшие уровни сознания

Мы уже анализировали проблему сознания для элементарного интеллекта. Был предложен механизм выделения и усиления одной модели в каждый момент при одновременном торможении всех остальных, осуществляемый с помощью гипотетической системы усиления-торможения (СУТ). Предполагается, что моделью в сознании может быть только короткая «фраза», что степень усиления и торможения определяется суммарным напряжением чувств, мерой которых является удовлетворение потребностей. Была рассмотрена система потребностей-чувств и особо выделены среди них рабочие, присущие каждому сложному интеллекту независимо от его назначения. Дано понятие о следящих системах, связанных со всеми ре-

цепторами и постоянно сигнализирующих о состоянии внешней среды и «тела». Их назначение — выдача «координат» сознания, то есть ориентировка интеллекта в отношениях со средой. Определена роль подсознания, в котором осуществляются простейшие ФА, не требующие напряжения, и готовится материал для сознания. Каждое переключение СУТ на новую модель является результатом активации ее со стороны предыдущих моделей, бывших в сознании, и всего комплекса моделей-координат.

Я повторил эти основные положения для того, чтобы было легче представить изменения сознания и подсознания в связи с усложнением интеллекта. Оно связано с увеличением объема памяти, а с ним — и количества моделей в активном состоянии, с возрастанием значения обобщения и, самое главное, с появлением речи. Учитывая присущее элементам интеллекта качество тренируемости, общество повышает его возможности самоорганизации. «Чувственное значение» моделей-образов и особенно моделей слов значительно изменяет критерии как источники активности моделей, а следовательно, изменяет направление мышления и поведения, то есть выбор ФА.

Мне кажется целесообразным выделить три условных уровня сознания как показатели совершенствования интеллекта. Простое возрастание количества моделей в памяти не отражает «уровень ума», хотя и является неременным его условием.

Первый уровень соответствует интеллекту ребенка или высших животных. Он предусматривает действие самого механизма СУТ, то есть выделения в каждый данный момент одной наиболее значимой модели и ее усиление с одновременным торможением других. По сути это механизм внимания. Существует слежение за внешней средой, «телом» и состоянием органов движения, действующее на подсознание и выполняющее сигнальные функции. Деятельность интеллекта осуществляется по ФА с неизменным числом потребностей-критериев, но с возможностью их тренировки и детренировки. Память обновляется обучением и забыванием, по законам тренировки элементов и связей. Ее возможности ограничены. Ограничена также способность к созданию обобщенных моделей; речь — только самая примитивная.

Второй уровень интеллекта связан с развитием речи и памяти. Появилась «вторая сигнальная система», 150

по И. П. Павлову, или искусственный экономный код для моделей образов, расширивший возможности общения интеллектов, а следовательно, возможности обучения и воспитания как средства изменения потребностей-критериев. Речь привела к счету, который сильно изменил восприятие мира. (Попробуйте сосчитать предметы, не пользуясь словами-цифрами!) Многие обобщенные понятия, трудные в образном представлении, определены словами речи, и обученный интеллект получил возможность выделить соответствующие образные модели. Сформированы понятия времени (настоящее, прошлое, будущее), реального и нереального, понятие «Я», вторых и третьих лиц, усовершенствовалось слежение за собственными действиями. Все это было заложено еще на первом уровне и использовалось интеллектом без речи, но «не осознавалось» и, следовательно, было неуправляемым. Так, на первом уровне «будущее» представлено как автоматическая программа прогнозирования, реальные образы отличаются от воспоминаний по их связям с задействованными рецепторами. Представлена также одна из основных программ — выделение «Я». Это необходимо, поскольку потребности «тела» неизбежно присутствуют даже в узкоспециализированных искусственных интеллектах. То же касается и слежения за собственными действиями — оно входит в самые простые ФА в качестве обратной связи, хотя и не осознается. Главное отличие второго уровня состоит в появлении убеждений в виде очень активных словесных формул, приобретающих значимость врожденных потребностей. (На этом у примитивных народов основаны всевозможные табу). Счет времени и развитая память удлинители предвидение — прогнозирование и породили возможность обобщенных ФА. Планы высших этажей, их модели приобрели чувственное значение — по тем же принципам, что и словесные формулы убеждений за счет связей с чувствами и вследствие многократного повторения.

Главное отличие сознания третьего уровня — это проникновение интеллекта в самого себя. Важнейшая черта — слежение за мыслями. Создается своеобразный «рецептор», наблюдающий за переключениями СУТ и отмечающий в памяти, к каким моделям оно приключается (иными словами, отмечающий, о чем интеллект думает). На более низких уровнях сознания самонаблюдение ограничивается наблюдением за

своими действиями, но не мыслями. Поскольку у человека на данном уровне развития обязательно имеются убеждения, иными словами, он «знает», что правильно, что неправильно, он имеет возможность периодически анализировать свои мысли, сравнивая их с моделями-эталоном, и в некоторой степени управлять ими. Такое управление связано с волей.

О «свободе воли». В психологии часто обсуждается этот термин. Будто бы именно наличием «свободы воли» отличается человек от животных: он может действовать произвольно, а животные только реализуют инстинкты. Боюсь, что это заблуждение. Когда «свободу воли» примитивно понимают по схеме «хочу — сделаю так, хочу — этак», то остается совершенно неясной природа этого «хочу». Анализируя механизмы мышления, его алгоритмы, я старался показать, что мысли и мышление — процесс детерминированный. Возникновение каждой новой мысли, а следовательно, и того самого «хочу» (что, между прочим, всегда выражается словами, а не только образами) предопределено взаимодействием моделей, часть которых вызвана к активности внешними воздействиями, а другая часть хранится в памяти как результат прошлой деятельности. Никакие «хочу» не появляются из ничего. Иллюзия «свободы воли» связана со сложностью анализа мышления. Одной из целей создания искусственного интеллекта как раз и является разоблачение ее.

Чаще всего воля проявляется в настойчивом выполнении принятого решения вопреки помехам. Она как бы дополняет «рефлекс цели» на уровне убеждений. Человеку внушают, что принятое решение надо осуществлять до конца, что отступать от этого правила плохо даже в том случае, если решение известно только самому себе. Все это типичный пример убеждений. Механизм воли — сильное возбуждение (высокая активность) слов, выражающих волю: «надо», «должен» и им подобных. Их активность должна превышать «тормозы» в виде утомления или других чувств — как биологических, так и социальных. Повторяю: воля — это убеждение, дублирующее «рефлекс цели» и действующее по тому же принципу: слежение за выполнением ФА и доведение его до конца.

Управление мыслями значительно сложнее, чем действиями. Особенности механизма сознания — переключения СУТ — таковы, что в нем нет достаточно сильной обратной связи, способной сигнализировать

о том, что мысли «ушли» с избранного направления. Здесь нужна высокая активность следящей системы, которая контролирует мысли. Добиться этого можно только тренировкой сосредоточения, специальными приемами, пример которых представляет йога. Максимально исключив сильные внешние раздражители и затормозив убеждениями биологические потребности, йог тренирует свою систему слежения за мыслями, чем достигает ее высокой эффективности. Активность моделей соответствующих слов становится такой, что, направив ее на любое действие, можно противостоять значительным «тормозам». Выбор направления как будто подчиняется «свободе воли», а в действительности обусловлен воспитанием йога, его убеждениями и памятью.

Третий уровень сознания совсем не обязательно предопределяет высокое развитие интеллекта в смысле обилия моделей в памяти и использования их для творчества. Примером являются те же йоги, которые, располагая всей мощью управления мыслями, тем не менее ничего не сделали для прогресса цивилизации. Однако способность к сосредоточенному мышлению является важным условием плодотворной деятельности интеллекта. Запоминание мыслей, хорошо представленное на этом уровне сознания, помогает в творческой работе.

Наблюдение за мыслями — главное в изучении самого себя и прежде всего своих чувств. Создается модель себя, после чего ее можно сравнивать с моделями, предлагаемыми обществом; одна из них принимается как образец, обычно через подражание авторитету. Однако не всегда дело обстоит так просто. Идеалы не только заимствуются, но и создаются заново. Это одна из сторон интеллектуальной деятельности.

Сопереживание и воображение

Остановимся вначале на явлении перевоплощения в личность другого человека — явлении сопереживания. Оно очень важно, так как представляет собой моделирование партнера по общению, необходимое для планирования поведения. Перевоплощение возможно только при хорошем развитии речи,

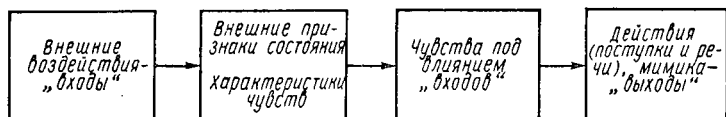


Рис. 38. Цепочка модели личности.

когда имеются словесные эквиваленты для обозначения состояний человека, воздействий на него, его чувств и действий. Пока речь будет идти только о человеке. Далее я попытаюсь показать, что все сказанное можно распространить и на искусственный интеллект.

Человек познает других и самого себя по тем словесным моделям, которым его обучает общество. Оно формирует модель личности, состоящую из элементов, показанных на рис. 38.

Возможно, что ребенок сначала создает самую простую модель близких людей, состоящую из двух элементов: «вход» — «выход». Потом он изучает внешнее проявление состояний и приписываемые им чувства, о которых ему рассказывают. На себя самого он переносит эти сведения потом — просто потому, что собственную внешность наблюдать труднее, чем внешность окружающих. Так постепенно модель личности усложняется, и ее главным звеном становятся чувства.

Исходя из общей формулы личности и наблюдая многих людей, человек делит их на типы: по социальному статусу, возрасту, полу, интеллекту и, самое главное, по предполагаемым характеристикам чувств (добрые, злые, жадные и т. д.). Чем богаче запас подобных сведений в памяти, тем большая детализация. В этом человеку значительно помогает знакомство с художественной литературой.

Модель личности применяется к партнерам по общению и позволяет рассчитывать их ответную реакцию на свои воздействия. Подобный расчет вполне алгоритмизируем, иными словами, он поддается расчленению на элементарные действия с моделями, как и всякое планирование ФА. Однако с перевоплощением дело обстоит сложнее. В тех случаях, когда оно подлинное, человек как бы испытывает чувства своего партнера, ставя себя на его место. При этом имеет место явление (назовем его «программой»), обозначаемое в психологии термином «воображение».

Рассмотрим воображение как ФА. С одной стороны, оно примыкает к планированию, элементу всякого

ФА, с другой — переходит в мечты или планы, лишённые реальности. Как известно, планирование — это построение моделей собственных действий, вызванных ими изменений объекта и предвидение своих чувств. В простом варианте предполагалось, что ФА направлен на удовлетворение потребности за счет «платы» от среды, полученной в результате действий, достигающих цели. Главным стимулом всех этапов ФА является приращение интегрального чувства *Пр* или уменьшение *НПр*, которое возможно только в том случае, когда есть неудовлетворенная потребность. Дополнительным стимулом являются рабочие чувства — любознательность, удовольствие, получаемое от выполнения дела, «рефлекс цели», «рефлекс свободы», а главным «тормозом» — утомление или скука от однообразия. К этому прибавляется еще адаптация, когда при повторном и легком удовлетворении потребности увеличивают притязания на «плату». Если среда не представляет сопротивления для удовлетворения потребности, то именно адаптация при повторном действии уменьшает средний уровень *Пр*, то есть понижает уровень душевного комфорта УДК.

Программа поиска в реальности или в мечтах — вот выход из скуки при сереньком благополучии. Она является функцией от рабочих критериев — любознательности, потребностей в информации, деятельности, преодолении трудностей. Эти потребности исключительно важны и присущи каждому сколько-нибудь развитому интеллекту. Они выступают на первый план тогда, когда насущная потребность удовлетворена, расширить притязания не позволяет среда, а основное чувство притупилось вследствие адаптации.

Суть поиска как самостоятельного вида деятельности состоит в передвижении в пространстве за пределы известного. «Искать новые места», «увидеть новое» — вот самая обобщенная модель действия поиска. Вторая общая модель — изменить деятельность, найти другой источник приятного. Если первая — это поиск «вширь», то вторая — поиск «вглубь».

Поиск имеет отношение к воображению как специфическому действию с моделями. У человека с высоким уровнем сознания это действие можно включить произвольно.

Воображение как мечта — это заведомо нереальное планирование. Как уже говорилось, есть чувство реального. Реальна модель внешней среды, когда в

ней присутствует «буква рецептов», указывающая на то, что модель объекта действительно является следствием его восприятия. Это относится ко всем временам («я действительно вижу», «видел», «делаю», «делал»). Как только настоящее прогнозируется в будущее, «буква» реальности утрачивает определенность. Будущее реально настолько, насколько вероятно предполагаемое событие и насколько оно удалено во времени от данного момента. Эти тонкости (с одной стороны, будущее не есть действительное, а с другой — оно в некоторой мере реально) учитываются специальным чувством реальности и отражены в речи различными вариантами сослагательного наклонения. Повторяю, мечты — нереальное планирование. Во всех других отношениях оно осуществляется по тем же алгоритмам: модель исходной ситуации оценивается, то есть определяются неудовлетворенные потребности и, следовательно, приращение чувств ($\Delta\mathcal{C}$) от возможной «платы». Это является стимулом для планирования, которое начинается как реальное. Однако сразу же выясняется отсутствие необходимых обстоятельств — как в объектах, так и в собственных возможностях. У совершенно реалистичного интеллекта ФА на этом прекращается. Чувство остается неудовлетворенным. Но мечтатель с этого пункта начинает мечтать.

Суть нереального планирования состоит в том, что к реальной исходной ситуации прибавляются заведомо или почти заведомо несуществующие обстоятельства. В такой роли выступают прежде всего действия, превышающие собственные силы. Либо же это могут быть внезапные изменения ситуации, появление недостающих деталей в объекте или новых объектов, которые изменяют уровень «ценности» и доступности среды для действий. Эти варианты подсказывает память, в которой хранятся «фразы» для успешных действий. Так или иначе, произвольное изменение начальных условий делает весь ФА успешным. Само воображение, то есть планирование, может остановиться на уровне наиболее обобщенного плана («Если бы я был умнее, я бы решил эту задачу и прославился») и т. п. Но настоящий фантазер не удовлетворяется короткой обобщенной фразой, за которой следует вздох «Но этого нет...», и начинает следующий этап планирования — разработку детальных планов, исходя из принятых допущений. Продумываются этапы действий, 156

соответствующие изменения объектов и сопутствующие чувства. Выбираются наилучшие варианты. Для всей этой работы нужны стимулы, поскольку необходима энергия, чтобы вызвать из памяти новые модели и оперировать ими. Стимулом является все та же неудовлетворенная потребность. Реальное планирование повышает вероятность получения «платы» по мере уточнения плана и тем самым увеличивает стимул или, по крайней мере, поддерживает его для противодействия «тормозу» — утомлению. Воображение или мечты странным образом «символически» удовлетворяют потребность, конечно, лишь в незначительной степени, однако достаточной для того, чтобы дать немного энергии, нужной для мечтаний. Немного, потому что выбираются модели, довольно хорошо известные, не требующие большого поиска в глубинах памяти. Таким образом, воображение базируется на способности моделей с «буквой» «нереально» активировать соответствующие потребности и даже представлять для них некоторую «плату». Это своего рода сопереживание самому себе.

Труд и творчество

В обществе принято считать трудом такие действия, результаты которых используются другими людьми, иными словами, в это понятие фактически входит только общественный труд. Это положение можно переложить на кибернетический язык таким образом: любой ФА направлен на внешнюю среду, служит удовлетворению потребностей, и, следовательно, его можно рассматривать как труд. С развитием общества усложнилось понятие «среда» и усложнились потребности. В этом все дело.

Алгоритм «трудовых» ФА ничем не отличается от алгоритма других, «личных» ФА. Те же этапы, тот же принцип иерархии и сети ФА с разной степенью общенности, тот же перерасчет стимулов. Так же различно напряжение, зависящее от чувств и «сопротивления» среды. Но это только в том случае, если в памяти есть достаточное количество моделей для всех этапов ФА — анализа, планирования, действий, то есть тогда, когда производится рутинная работа.

Если же готовых моделей нет, требуется творчество.

157 Творчество — это создание новых моделей и воплощение их через ФА в материальные вещи или в ма-

териальные модели — книги, рисунки и т. д. Творчество постоянно присутствует в планировании любого ФА. Оно наличествует в небольшой степени даже у животных, поскольку и они имеют набор пробных движений, из которых «монтируются» необычные ФА. Просто у человека неизмеримо сложнее и длительнее планирование.

Любая новая модель создается из известных элементов. Элементами могут служить модели разной сложности. Для каждого сложного объекта можно предложить бесконечное множество моделей. Так же бесконечно и творчество.

Где же критерий ценности моделей, создаваемых в результате творчества?

Этот критерий — в практике, в применении модели для удовлетворения определенных потребностей.

Есть различные виды творчества, они требуют разной методики проверки ценности его результатов.

На первое место поставим изобретение вещей — искусственных материальных объектов. Вопрос об их полезности решается просто.

Следующий вид творчества — это наука, создание моделей сложных объектов. В основе его лежат наблюдения структуры и функций объекта или восприятие уже известных его моделей. По ним выдвигаются гипотезы — наборы «фраз», в которых наряду со «словами» известного присутствуют «слова» предполагаемого. Выдвижение гипотезы — часть этапа ФА, который я назвал анализом. Гипотезы выдвигаются при любом исследовании объекта, думаю, они возможны даже у животных. Весь вопрос — в сложности задачи, в соотношении «слов» известных и неизвестных. Правомерность гипотезы проверяется последующими целенаправленными исследованиями, а если это невозможно, то хотя бы тем, что гипотезу принимают другие интеллекты. Это, разумеется, еще не делает ее истиной, но свидетельствует о том, что она соответствует уровню времени.

Еще один вид творчества — искусство. Но о нем я скажу немного позднее.

Алгоритм творчества в самом общем виде представляет собой развертывание обобщенных моделей в детальные. Схема этого процесса показана на рис. 39.

Восприятие объекта дает модель — в той или иной степени обобщения. Она оценивается по чувствам, планируется действие и создается модель цели, как пра-

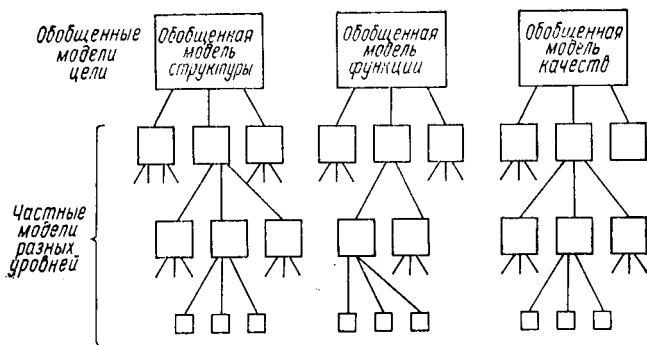


Рис. 39. Схема развертывания обобщенных моделей структуры, функций и качеств объекта цели в детальные для создания непротиворечивой модели.

вило, тоже в обобщенном виде. На рисунке показана обобщенная модель цели в трех отдельных выражениях — в моделях структуры, функции, качеств. Степень обобщения может быть очень различной, код моделей тоже. Неодинакова и значимость каждого из трех типов моделей — для одних важно качество, для других — функция, для третьих — структура. По общим правилам действия с моделями обобщенные модели могут быть развернуты в некоторое множество более детальных моделей — в зависимости от параметра, по которому произведено или предусмотрено обобщение. Количество частных моделей, входящих в множество, объединенное в обобщенной модели, различно и зависит от знаний интеллекта, то есть объема его памяти. На рис. 39 показана иерархия обобщений в четырех уровнях. Число таких уровней может быть велико, состав каждого весьма разнообразен. Например, нужно изобрести машину, для которой известна только ее обобщенная функция: допустим, она должна что-то подбирать и передвигать. Структура не задана, но заданы качества — в крайних пределах размеров, веса, прочности и пр. По таким обобщенным моделям функций и качеств можно вообразить бесконечное множество не только конструкций, но и самих принципов действия машины — механических, электродинамических, пневматических, гидравлических и пр. В одном случае главной моделью является обобщенная структура (к примеру, в архитектуре — башня), а другие обобщенные модели — функций и качеств — имеют

подсобное значение. Высшее обобщение качества можно развернуть в очень большую серию обобщений по признакам — цвет, цена, тяжесть и т. д. Каждый в свою очередь можно еще детализировать, и так до тонкостей.

Обобщенные модели цели не представляют никакой ценности, и их может создать каждый («изобрести бы машину для того-то» и т. п.). Творчество состоит в развертывании обобщенных моделей цели в детальные, причем так, чтобы они удовлетворяли главному критерию ценности, о котором говорилось. Как правило, творец действует методом перебора: спускаясь по ступеням обобщения с верхнего уровня на следующий, он перебирает уже имеющиеся в его памяти модели. Обычно перебор частных моделей начинается с главной обобщенной — со структуры, функций или качеств. Каждую выбранную частную модель проверяют по другим параметрам обобщения, при этом сразу же выбраковывается абсолютное большинство вариантов. Каждая из оставшихся моделей, которая удовлетворяет требованиям обобщения второго уровня, развертывается в более детальную модель третьего уровня и т. д., после чего снова идет проверка по другим критериям, уже детализированным до соответствующего уровня. Общий алгоритм творчества (как и планирования) хорошо представлен в конструкторском труде.

Казалось бы, раз известен такой алгоритм творчества, отчего не создавать гениальных произведений? Перебирай и перебирай, подключи к этому компьютер, если мозг работает медленно.

Есть несколько препятствий на пути к успеху творчества. Главное — великое множество деталей, которые, последовательно проходя через уровни обобщения, достигают вершины в общей модели цели. Это множество невозможно удержать в памяти, невозможно перебрать, поскольку оно включает в себя значительную часть всех моделей, накопленных человечеством. Нечего и говорить, что перебор неосуществим, если начинать «снизу» — с малых деталей. Даже опускаясь «сверху», последовательно отбрасывая целые крупные группы моделей, исключительно трудно получить необходимую структуру, если учесть, что нужно подобрать не только модели-элементы, но еще правильно их соотнести друг с другом. При этом неизвестно даже число элементов. Нет, расположение моделей по иерархии обобщения и использование алгоритма перебора

отноудь не обеспечивают легкого достижения цели.

Вторая трудность — в отсутствии необходимых деталей или, скажем, частных моделей — структурных подробностей. Их накоплено очень много и тем не менее совсем недостаточно для того, чтобы решить любую научную или техническую задачу. В познании сложного объекта одинаково важны как обобщенные модели, так и сугубо частные детали. Например, в развитии биологии огромную роль сыграло открытие того, каким образом из нуклеотидов складывается двойная спираль ДНК. Все широкие гипотезы о жизни без таких деталей не позволяют довести модели до высшего подтверждения их истинности — создания новых живых существ. Эта задача еще не решена, но подходы уже вырисовываются, например создан живой ген.

Детали добываются так же трудно, как и обобщения. Это черновая работа науки и производства.

Третья трудность — отсутствие четкости в обобщениях моделей. Это только в схеме существует иерархия обобщенности: крупные блоки, детали и пр. В действительных моделях процесс обобщения нестрого формализован и уровни обобщенности не существуют как нечто стабильное. По крайней мере, так обстоит дело в естественном сетевом интеллекте, каким является кора мозга. Правда, при создании ИИ можно задать дискретные уровни обобщения, и они помогут при переборе вариантов, но полностью не разрешат проблемы оптимума.

Переход от обобщенной модели к более детальной осуществляется по связям, которыми сформирована короткая «фраза»: обобщенная модель → частная модель. Следовательно, в памяти алгоритмического интеллекта нужно иметь множество «словарей», фиксирующих обобщения. По ним будет осуществляться перебор-поиск частных моделей по обобщенной. В сетевом интеллекте коры мозга обобщенная модель представляется просто неясной частной моделью и связанной с ней моделью-«буквой», обозначающей действие обобщения, то есть последовательного сравнения ряда частных моделей по некоторому признаку. Пример — образ «четвероногое». Здесь обобщение произведено по признаку «четвероноготь» на серии конкретных моделей четвероногих животных. Модель «четвероногое», как таковая, существует только в виде слова речи. Слово соединено связями различной проходимости с конкретными моделями — образами четвероно-

гих животных. Такие модели имеют неодинаковую активность. Поэтому при перечислении животных с четырьмя ногами первыми вспоминаются наиболее знакомые, а затем — другие в порядке убывания проходимости связей и активности частных моделей, зависящих от частоты их использования.

Как говорилось, СУТ в сетевом интеллекте тормозит модели. Поэтому вспоминание частной модели по обобщенной, как действие перебора, иногда происходит не сразу: модели-адресаты могут оказаться сильно заторможенными, и вспомнить их не удастся до тех пор, пока они не освободятся от тормозящего действия СУТ. Всем знакома эта особенность вспоминания: знаешь, что в памяти был образ или слово, иначе говоря, к нему есть линии связей, а вспомнить не можешь. Проходит время, и нужное слово всплывает в сознании само собой.

Организация памяти в алгоритмическом интеллекте — самая трудная проблема. Признаков-параметров, по которым можно обобщать модели, великое множество. Пример — те же «четвероногие». Значит, нужно много «словарей». Кроме того, внутри «словаря» необходим порядок значимости моделей, отражающий частоту их использования. Этот порядок должен периодически пересматриваться в зависимости от повторного привлечения модели в сознание. Конечно, переборы во внешней памяти компьютера — дело трудоемкое и требующее машинного времени, но положение не безнадежно, поскольку пересмотр массивов памяти возможен в интервалах между основными действиями.

Простое перечисление трудностей перебора показывает, что успешное творчество тем менее вероятно, чем выше степень обобщенности задач и чем больше количество моделей, описывающих цель. Гениальные изобретения или смелые гипотезы отличаются тем, что их творец использует модели, очень далеко отстоящие от проторенных путей поиска, такие, которые не приходят в голову при простом переборе по порядку используемости или значимости.

Все слышали об «озарениях», когда решение трудной задачи приходит неожиданно, иногда в самое неподходящее время. В чем тут дело?

Исходя из принципов действия сетевого интеллекта, такие явления можно представить себе как поиск в подсознании. Не случайно гении много думают о задаче. Это значит, что исходные обобщенные модели,

которые всегда направляют поиск, у них очень активны, натренированы частым привлечением в сознание. Даже будучи приторможены со стороны СУТ, они передают активность по связям на все модели, имеющие отношение к проблеме. Вследствие этого активность вторичных моделей, связанных с главными, тоже возрастает. В подсознании все время светится «фраза» задания. Если в результате колебаний активности моделей при очень большом количестве связей «вдруг» всплывает новая модель, то она немедленно замыкается на «фразу» задания и привлекает СУТ. Это и есть знаменитая «Эврика!».

Таким образом, с одной стороны, благодаря работе СУТ активируется весь комплекс моделей, связанных с задачей, причем не обязательно, чтобы все они привлекались в сознание, а с другой — в подсознании происходит «игра активности» как процесс, в значительной мере случайный. Наслоение обоих факторов может с некоторой вероятностью дать нужный результат. При условии, конечно, что есть еще и знание, иными словами, наличие в памяти многих моделей. Вероятность удачной находки — с учетом случайности — невелика, но ведь и «озарения» случаются крайне редко. Как правило, нормальное творчество идет все-таки методом сознательного перебора — от обобщенных моделей к частным. Проверка вариантов по многочисленным критериям производится человеком при помощи его «внешней» памяти — собственных описаний, схем и расчетов, зафиксированных на бумаге. Эти модели потом повторно воспринимаются через сознание наравне с чужими моделями и в свою очередь дают новый толчок активности всем моделям, имеющим отношение к проблеме.

Возможно ли воспроизвести в алгоритмическом интеллекте предположенный для человеческого разума механизм творчества? Иначе говоря, возможно ли осуществить в подсознании часть работы по перебору вариантов? Мне представляется, что да, хотя и не в полной мере. Все упирается в организацию памяти: соотношение моделей разной обобщенности, их связи во «фразы». Можно создать подпрограммы, которые будут осуществлять свою работу поиска независимо от сознания.

Творчество в искусстве имеет некоторые особенности. Произведение должно быть умно, красиво и не оставлять равнодушным. Стимулы для творчества

такие же, как и для любого ФА: специальные и рабочие — удовольствие от совершения дела и новизны. Важной является потребность самовыражения через общение.

«Умно» означает представление автором сложных моделей, отражающих меру глубины его понимания сюжета. Для того чтобы произведение не оставляло равнодушным, оно должно содержать подбор слов, форм, красок и т. д., способных активировать чувства по тем принципам, которые уже были описаны. Разумеется, для этого у человека, воспринимающего произведение искусства, должны уже быть связи между соответствующими моделями и чувствами, иначе оно не будет понято. Однако «язык чувств» врожденный, и это облегчает задачу. Гораздо сложнее объяснить, что такое «красиво». Видимо, для эстетического чувства нет биологического аналога, оно — целиком производное общественного развития человека. Это доказывается общеизвестным фактом изменения стандартов красоты в различные исторические эпохи и у разных народов. Неизменным является только любовь к красоте природы, для этого, возможно, есть биологические корни. Тепло, солнце, зелень, вода — все это сочеталось с приятными чувствами и, наверное, закрепилось в генах. Словесные и графические модели природы тоже приятны, воздействуют на те же чувства.

Можно ли алгоритмизировать художественное творчество? На сетевом интеллекте принципиально можно, но практически до этого очень далеко, так как пока не предвидится технологии создания СИ большой мощности. Возможности алгоритмического интеллекта ограничены. В частности, это выражается в отсутствии такого непосредственного взаимодействия между моделями в кратковременной памяти, как это имеет место в сетях. Следовательно, чувства беднее и выражение их словами труднее.

Сновидения и психологические болезни

Человек и животное видят сны. Иногда это предельно простые видения — образы, иногда сложные повествования, в которых воспроизводятся такие ситуации, которые никогда не происходили в реальной жизни. Во сне испытываешь яркие чувства

и их крайнее выражение — эмоции радости, ужаса, гнева. Во сне можно даже сделать изобретение.

Гипотеза о механизмах мышления должна объяснять, а модель интеллекта воспроизводить феномен сновидений. Мне кажется, что это возможно.

У нас есть кратковременная память возбужденных моделей, в которой после уменьшения их активности проходимость связей еще сохранена на довольно высоком уровне. В этой памяти масса коротких «фраз», состоящих из нескольких «слов» — моделей различной степени обобщенности, объединенных в ансамбли связями. Одному сложному образу или ситуации соответствует комплекс таких «фраз», не очень прочно связанных друг с другом. Наиболее яркими являются модели-образы, они представлены отдельными картинами и их последовательностями. Во время бодрствования все картины имеют «букву» реальности — привязку их к моделям настройки рецепторов, воспринимавших эти картины, и «буквам», обозначающим время. Однако связи между «координатами» сознания не столь прочны, как между отдельными ансамблями образов, и картины реального бытия удерживаются в состоянии целостности благодаря действию СУТ. Когда одна из моделей попадает в сознание, она получает дополнительную энергию, которой достаточно для того, чтобы пробивать плохо проторенные связи и возбуждать последовательность моделей, в том числе и моделей-«координат». Главной среди них является обобщенная модель бытия, сознания — чувство реальности, составленная из всего комплекса рецепторов. Таким образом, деятельность СУТ и «координата» реальности организуют весь «поток мышления». При достаточно высоком уровне сознания можно произвольно включить нереальные действия с моделями — мечты.

Представим себе, что СУТ выключена. Активность моделей в кратковременной памяти сохраняется на низком уровне, поскольку они были возбуждены в предшествовавшее время и взаимодействуют друг с другом. Как было указано при обсуждении деятельности мозга, нейроны, составляющие модель, тесно связаны друг с другом и поддерживают активность модели-ансамбля как единого целого. Она обладает собственной активностью, зависящей от уровня тренированности нейронов. Модели-ансамбли взаимно возбуждают и тормозят друг друга, так что уровень их активности постоянно колеблется. При этом возможен

некоторый порядок — последовательное выделение наиболее активных моделей при значительном торможении других. Выше даже высказывалось предположение об отсутствии в мозге специальной системы СУТ, а выделение одной главной модели в каждый момент — механизм сознания — объяснялся законами взаимодействия моделей при общем активирующем влиянии подкорки. Конечно, это возможно только в СИ, в АИ для воспроизведения сознания СУТ необходима.

Сновидения в естественном мозге можно объяснить сильным уменьшением активирующего влияния со стороны подкорки, в результате чего нарушается самоорганизация «потока мышления» с четкой привязкой моделей к «координатам» времени, пространства, самонаблюдения. При полном выключении подкорковой активности механизмы доминирования не функционируют, и продолжается лишь беспорядочное несильное возбуждение моделей за счет их остаточной активности. Однако если механизм центральной активации подавлен не полностью, но рецепторы тем не менее отключены, то доминирование среди моделей кратковременной памяти продолжается и действует некий суррогат сознания: активность циркулирует только по связям между моделями, которые были активированы в наибольшей степени в предшествовавший период бодрствования. «Боковые цепи» сознания — «координаты» времени и пространства, а также модели речи обычно остаются заторможенными, как и центр реальности. Фактически действует подсознание. Механизм доминирования еще сохраняется, следовательно, сохраняется движение «мысли», то есть переключение активности по цепочке моделей, которое и создает подобие связного мышления.

Сновидения напоминают мечты и воображение: так же живо представляются картины невозможных действий, иногда самого фантастического свойства. Разница в том, что при мечтании осознается полная нереальность происходящего. Именно поэтому воображаемые картины действуют на чувства слабее, чем сновидения (вспомним выражение: «Проснулся в холодном поту от ужаса»). Речь представлена в сновидениях наиболее бедно. Наверное, это происходит потому, что внутренняя речь — заторможенное мышечное действие, а оно требует участия обратной связи — рецепторов в мышцах.

Еще один вопрос — творчество во сне. Механизмы творчества в сетевых моделях мы уже рассмотрели. Было показано, что создание новых моделей — «фраз» — может происходить в подсознании в тех случаях, когда существуют очень активные моделизадачи, генерирующие энергию на другие модели, пока не произойдет «замыкание» на нужную модель. Этот механизм вполне может действовать и во сне. «Эврика!» бывает настолько важна, что человек просыпается.

Моделирование сновидений на моделях сетевого интеллекта мне кажется вполне реальным. Нужно постепенно уменьшать активность СУТ и тормозить настройку всех рецепторов вплоть до полного прекращения восприятия внешнего мира. Отключение внешних рецепторов автоматически уменьшает чувство реальности, и интеллект как бы оказывается вне времени и пространства. Торможение рецепторов «тела» еще более усиливает это положение.

Воспроизведение снов в АИ едва ли имеет смысл, поскольку интеллект такого типа не предназначен для доказательства гипотезы о механизмах деятельности мозга. Он призван воспроизвести гипотезу о сущности интеллекта более высокого уровня обобщенности, тогда как сновидения — только побочный продукт функции естественного сетевого интеллекта.

Тем не менее в порядке исследования АИ можно попытаться создать условия физиологического сна. Возможно, сновидения будут воспроизведены. В самом деле, если отключить все следящие системы — за внешней средой, «телом», мыслями, то ощущение реальности исчезнет. Если при этом сильно понизить характеристику СУТ, то возрастет роль подсознания и уменьшится роль сознания. Возможно, организуется поток картин вне связи с «координатами» времени и пространства, и это будут сны.

Патология искусственного интеллекта — совсем другое дело. Если он «сойдет с ума», то может натворить немало бед, поэтому исследование такой возможности имеет большое значение.

Патологическое поведение человека характерно отличием от общепринятых критериев, нормы. Но критерии — это чувства. Разнообразие их характеристик придает людям индивидуальность. Давно известно, что нет четкой грани между сумасшедшим и психически здоровым. Больной отличается от здравомыслящего

при обобщении (интегрировании) его поступков во времени, но и здоровые люди иногда совершают неразумные действия.

Каждый акт мышления и поведения таит в себе возможность патологии. Наша гипотеза позволяет дать анализ многим психическим заболеваниям, но не стоит с этим спешить, вначале следует показать на моделях нормальное поведение. Поэтому я ограничусь кратким экскурсом в сферу психопатологии.

Понятие «галлюцинация» обозначает нереальное восприятие, когда больной принимает воображаемое за действительность. Видимо, это можно воспроизвести, изменяя характеристики чувства реальности, «координаты» пространства и настройки рецептора. При этом реальные пространственные картины будут перемежаться воображаемыми, и грань между ними может исчезнуть.

Неразумные поступки чаще всего проистекают от изменения оценок правильно воспринятых внешних воздействий. Оценки строятся по характеристикам центров потребностей-чувств. Если их сместить в ту или другую сторону — горькое будет казаться сладким, плохое — хорошим или наоборот. За этим последуют соответствующие неправильные действия. Когда такое смещение касается одного чувства, тогда мы имеем дело со «странностью» (скажем, человек начинает бояться пространства и не выходит на улицу). Так называемые пороки тоже иллюстрируют смещения чувств. Обжоры, сластолюбцы и честолюбцы не считаются больными, хотя их поведение нередко выходит за границы общепринятой нормы. Воспроизвести такие отклонения поведения на модели не представит особого труда.

Гораздо серьезнее смещение соотношения активности интегральных центров Pr — $НPr$. Если нулевая линия безразличия смещена кверху, то даже нормальная «плата» (то есть внешнее воздействие) оценивается как недостаточная, и, следовательно, человек ощущает неудовлетворенность, которая порождает ущемленность, подозрительность, эмоцию горя, может вызвать двигательную реакцию гнева или, наоборот, апатию. Слабое смещение нулевой линии характеристик кверху выражает пессимистический склад психики, сильное смещение свидетельствует о депрессии.

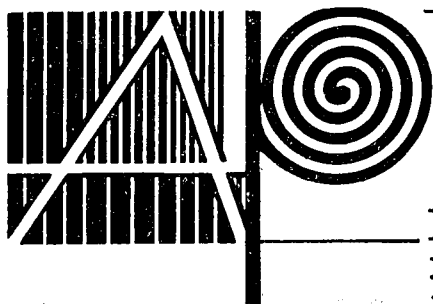
Противоположный сдвиг нулевой линии дает оптимистов, интегральное чувство Pr у них преобладает 16

и диктует соответствующее поведение: веселый нрав, эмоции радости при малейшем поводе.

Если совместить изменение активности чувств *Пр — НПр* с изменением характеристики СУТ, то можно получить картину, напоминающую заболевание, называемое маниакально-депрессивным психозом. Для депрессивной фазы повышение значимости *НПр* нужно сочетать с замедлением действия СУТ, с понижением ее эффекта активации и усилением торможения. Наоборот, повышение значимости *Пр* вместе с активацией усиливающего эффекта СУТ и понижением тормозящего даст маниакальную фазу того же психоза.

Изменения характеристики СУТ, отдельных частных потребностей, убеждений, значимости «чувства реальности», соотношения *Пр — НПр* в сочетании с изменениями параметров памяти позволяют получить бесконечное многообразие изменений поведения. Можно предвидеть очень интересные работы в этом плане, особенно если учесть, что психиатрия остается пока эмпирической наукой.

Исследование патологии на модели искусственного интеллекта представляется важным для практики. Если «психика» ИИ будет неустойчива, то потребуются специальные конструктивные решения, ограничивающие изменения характеристик рамками безопасности. Это имеет отношение к «законам робототехники» Азимова.



Искусственный интеллект в человеческом обществе

Сначала обратимся к искусственному интеллекту, создаваемому для управления сложными системами, либо с целью делать это лучше, чем люди, либо для работы в условиях, где люди не могут существовать. Можно предположить такие примеры:

1. Роботы с интеллектом для работ под водой и в космосе, особенно на Луне и планетах.

2. ИИ для управления технологией.

3. ИИ с комплексом диагностической и управляющей аппаратуры, предназначенный для лечения больных.

4. ИИ для обучения и даже воспитания детей. Физическое воплощение такого воспитателя еще требует проработки.

5. Наконец, ИИ для управления экономикой общества. Кроме первой задачи, все другие требуют очень высокого уровня интеллекта, во всяком случае выше среднего человеческого, но притом ИИ должен обладать еще некоторыми дополнительными качествами, иначе не будет смысла в его применении.

Основные «индивидуальные» качества ИИ

Для такого ИИ, живущего среди людей, прежде всего нужна речь в самом широком смысле этого слова — устная, письменная, понимание графических и математических моделей.

Необходимым качеством должна быть обучаемость, поскольку невозможно обеспечить разумное взаимодействие со сложными динамическими системами без программы постоянного изменения моделей и приспособляемости.

Вопрос о воспитуемости как втором проявлении принципа самоорганизации представляется более сложным, и мы его рассмотрим ниже. Предварительно скажу: воспитуемость, то есть способность изменять характеристики заданных изначально потребностей и чувств, нужна обязательно. Уточнения требует лишь ее степень. Человек и животные тоже воспитуемы, но в разных пределах.

Разнообразие людей выражается в разнице характеристик врожденных чувств и характера, а также в способностях. В частности, в особенностях элементов памяти.

Разнообразие интеллектов, даже одинаково запрограммированных, должно возникнуть само собой в результате самоорганизации в процессе деятельности.

Уровень сознания ИИ зависит от его назначения. Для подводного робота достаточно и второго уровня, для всех же других «профессий», если их выполнять лучше, чем люди средних возможностей, необходим третий уровень.

Врожденные потребности и убеждения являются основными источниками активности, управляющими всеми действиями с моделями. Интеллект всегда утилитарен, даже если он нацелен на совершенно отвлеченные идеи. В ИИ на физических элементах врожденные потребности закладываются в виде «центров» — моделей со специфическими характеристиками, работающими как генераторы активности. В АИ активность — лишь параметр, выражаемый цифрами, но через его изменения осуществляется все регулирование действий с моделями. Изменение «врожденных» характеристик в процессе деятельности — главное проявление воспитуемости. Другим проявлением является создание убеждений.

Видимо, для каждого искусственного интеллекта необходима по крайней мере одна специальная потребность — самосохранение. Сюда входят питание и безопасность. Первое проявляется во внешнем энергетическом снабжении, второе — не только в защите и стремлении избежать внешней угрозы, но и в «боли» — как сигнале о внутренних неполадках. Любая

разумная машина, рассчитанная на самостоятельное существование, должна «ощущать боль» при повреждениях и иметь системы надежности. Внутреннюю «боль» ИИ легко воспроизвести, но внешняя угроза имеет информационный характер и требует программы выделения и распознавания образов угрозы. Потребности самосохранения выражаются в соответствующих чувствах, трактуемых на человеческом языке как голод, боль и страх. Страх — более широкое понятие, чем чувство, вызванное угрозой, действующей в данный момент. Это слово выражает также и чувство от любых будущих воздействий, угрожающих удовлетворению любой потребности. «Боюсь голода, боли, скуки, старости» и «сейчас, в данный момент, боюсь собаки или хулигана» — это не одно и то же.

Рабочие потребности и соответствующие им чувства обязательны в искусственном интеллекте высокого ранга. Главная из них — потребность в новизне (или просто любознательность) — особенно важна. Если создавать ИИ-исследователь, то именно это чувство должно стать главным стимулом его поведения. Столь же интересны потребность в деятельности и «рефлекс цели» — они действуют при любом ФА, хотя в большинстве случаев не обеспечивают сколько-нибудь значительных стимулов и не могут создать высокого напряжения действия.

Универсальной является потребность в расслаблении и отдыхе как реакция на любую напряженную деятельность. Активность чувства утомления является главным «тормозом» и в значительной мере определяет дееспособность ИИ в любой сфере. Спрашивается, нужно ли вообще задавать это чувство для искусственного интеллекта, призванного общаться с людьми? Мне кажется, что оно выполняет полезные функции, регулируя ФА, определяя расход энергии стимулов. Если его не будет, возможен «разнос», чрезмерное усиление одного вида действий в ущерб другому (особенно, если задано качество тренируемости элементов ИИ).

Чувство скуки, ощущение «надоело» сродни утомлению, оно тоже является «тормозом» и ограничивает продолжительность однородной деятельности, требуя переключения, хотя и не связано с расходом энергии. Необходимость такого чувства очевидна, хотя чрезмерное развитие его сделает ИИ бездельником.

Потребность в свободе, или «рефлекс свободы», по И. П. Павлову, является источником агрессивности

у людей. Как упоминалось, суть его выражается в том, что при возникновении помехи, угрожающей выполнению действия, животное прекращает основное действие и пытается уничтожить помеху. Соответствующее чувство можно определить как раздражение, более сильная его степень — злоба, а действия, вызванные ими, принято называть агрессией. Этологи в последние десятилетия много занимались изучением агрессии, справедливо полагая, что она важна и для понимания человеческих отношений. Большинство ученых сходится во мнении, что агрессия животных — это защитная реакция против любых помех, препятствующих выполнению запрограммированных инстинктов. Весь вопрос — в количественных характеристиках этой потребности. Главной сферой проявления агрессии среди животных является внутривидовая борьба. Нам кажется, что самец бросается на другого самца без всякого повода, и это трактуется как качество злобности. В действительности же самцы — конкуренты, поскольку каждый из них запрограммирован природой как потенциальная помеха программе размножения другого, и агрессия тут биологически оправдана.

Как же быть с «рефлексом свободы» и агрессивностью в программах искусственного интеллекта, предназначенного для взаимодействия с людьми?

Агрессивность в широком понимании слова в связи с «рефлексом свободы», видимо, полезна для ИИ. Прежде всего по той причине, что она есть у людей. Если ИИ должен иметь с ними дело, то некоторая доля агрессивности ему необходима.

Создать алгоритм для реализации «рефлекса свободы» с небольшой долей агрессии не представляет труда. «Входами» могут быть любые помехи в выполнении ФА, «выходами» — новые ФА, направленные на помеху.

Теперь самое время сказать несколько слов о «характере» ИИ. Прежде всего это степень «лени». Иначе говоря, значимость утомления или скуки, ограничивающая величину и длительность любого напряжения, а следовательно, и способность преодолевать сопротивление среды. Второе — характеристики СУТ: насколько она усиливает выбранные модели и тормозит все остальные, находящиеся в кратковременной памяти. Имеет значение не только ее статика, но и динамика — длительность приключения СУТ к очередной модели-«мысли». Соотношение активации и торможения

от СУТ создает разные типы «характера»: «возбуди-
мые» или «тормозные», а способность к напряже-
нию — сильные и слабые, которые делятся еще на
«импульсивные» и «уравновешенные» в зависимости
от динамики напряжения. Разнообразие характеров
вместе с разной значимостью врожденных чувств опре-
деляет индивидуальность личности. Разумеется, «ха-
рактер», определяемый особенностями СУТ и степенью
«лени», необходим ИИ.

Я много занимался проблемой моделирования лич-
ности и установил значение «общественных» потреб-
ностей-чувств. У животных они составляют стадный
инстинкт. Без него невозможно представить себе су-
ществование высших млекопитающих. Стадный ин-
стинкт применительно к человеку признается не всеми
учеными. Многие склонны относить социальные чув-
ства людей целиком за счет воспитания. Я далек от
биологизации человека уже просто потому, что реали-
зация любого инстинкта у человека в значительной
мере управляется общественными институтами. Кроме
того, огромную роль в формировании чувств человека
играют убеждения, присущие только ему. И тем не
менее, существуют врожденные задатки для социаль-
ных чувств и социального поведения, неодинаковые
у разных людей, и их нельзя сбрасывать со счета.
Чтобы уменьшить нежелательные наследственные чер-
ты стадного инстинкта, следует просто усилить воспи-
тание. «Человек как человек» не является темой этой
книги, но проблемы создания искусственного интел-
лекта для использования его в человеческом обществе
не позволяют совсем обойти этот вопрос. Пусть чита-
тель воспринимает его как спорное отступление.

Общественные потребности и чувства ИИ

На первое место здесь нужно поста-
вить потребность в общении. У животных она обеспе-
чивает возможность выбора полового партнера и игра-
ет важную роль в эволюции. У человека она действу-
ет столь же сильно, как и голод: в период вынужден-
ного одиночества возникает непреодолимое желание
говорить с людьми или хотя бы видеть их. После не-
которого периода общения человек устает и склонен
побыть в одиночестве. Потом все повторяется снова.

Для ИИ потребность в общении необходима, так как она побуждает его к контактам с людьми, без которых он не сможет познавать человеческое общество. Книг для этого недостаточно. Они не могут дать моделей-образов. Реализация потребности в общении не трудна. На «входе» «центра общения» будут действия, зрительные и словесные контакты, «выходом» явятся ФА общения. Они будут тормозиться утомлением и скукой от однообразия.

Общение связано с потребностью самовыражения. О ней уже упоминалось: у животных позы и мимика выражают чувства и выполняют сигнальную функцию. У человека к этому прибавилась речь. Желание высказаться — стимул от этой потребности. При остром стремлении к общению человек жаждет хотя бы посмотреть на других людей, посмотрев, хочет послушать их, насытившись слушанием, должен высказаться, притом так, чтобы его слушали. Но это еще не конец цепочки, еще не все. Человеку свойственно стремление «командовать», навязывать свою волю ближнему. Это называют потребностью в лидерстве. Психологи отводят ей большое место в мотивах поведения людей. Соревнование и лидерство играют огромную роль в эволюции, обеспечивая улучшение вида при половом отборе.

Моделирование лидерства как потребности быть первым и навязывать свою волю другим связано с теми же трудностями, что и модельное воспроизведение любознательности: «входами» является сигнальная информация, полученная из внешней среды через дистантные рецепторы. Нужно увидеть соперника, определить его действия, сравнить со своими, вывести разницу и это подать на «вход» «центра лидерства». «Выход» будет в виде стимула, который может усиливать обычную деятельность, с тем чтобы обогнать соперника или стать причиной агрессии. Таким образом, лидерство — типичный пример информационного критерия, когда деятельность определяется не «телом», а качествами внешней среды, избирательно представленной живыми существами данного вида.

Нужно ли лидерство искусственному интеллекту в тех сферах деятельности среди людей, которые были перечислены выше? Мне кажется, что да, хотя бы потому, что оно есть у людей. Иерархия положений — не формальная по признакам богатства или власти, зависящая от социальной и экономической структуры

общества, а та, что создается в любом коллективе, делающем общее дело, в значительной степени стимулируется лидерством. Оно — основа всякого авторитета, поскольку его приходится завоевывать. Одних способностей, знаний здесь недостаточно. Даже сильный характер, который часто сочетается с лидерством, должен получать дополнительный стимул от соревнования. Разумеется, чрезмерное стремление к лидерству — порок. Оно выражается в тщеславии и честолюбии, в эгоизме и даже жестокости. Следовательно, «центру лидерства» ИИ нужно задать соответствующие умеренные характеристики и ограничить их воспитуемость. Тогда будет исключена вероятность превращения его в диктатора.

Есть в психике животных и человека потребность — чувство, противоположное лидерству. Я назвал его подчиненностью. Оно примерно соответствует понятию «конформизм». Суть его — в склонности следовать за лидером, подражать ему, подчиняться авторитету. При этом человек испытывает удовольствие. Качество подчиненности присуще в той или иной степени всем, и, видимо, уровень его находится в обратной зависимости от уровня лидерства, которым в разной степени тоже обладают все. В модельном плане механизм этой потребности — подчиняться высшему — такой же, как и механизм лидерства: результат сравнения соперника с собой вляется «входом» для соответствующего «центра». Весь вопрос — в характеристике «центров»: человек с сильным стремлением к лидерству откажется от соревнования и ограничится вторым местом только в том случае, если соперник обладает очень большим преимуществом; человек же со слабо выраженным стремлением к лидерству легко сдается и даже получает удовольствие от подчинения. Наблюдения над людьми все это легко подтверждают: прирожденный лидер обо всем судит самостоятельно, даже если ничего не понимает в предмете, подчиненный признает авторитетом человека, едва возвышающегося над его собственным уровнем.

Моделирование подчиненности производится по тем же принципам, что и моделирование лидерства. Видимо, оно нужно для искусственного интеллекта хотя бы на период обучения. Без авторитетов трудно представить себе создание убеждений при высоком уровне сознания, когда способности к анализу поведения самого себя и окружающих развиты хорошо.

Есть еще одно чувство, производное от стадного инстинкта, — это сопереживание (проще — доброта или, строже, альтруизм). Разумеется, оно должно быть присуще искусственному интеллекту, живущему среди людей. Я уже описал механизмы перевоплощения и сопереживания. Они позволяют понять психику другого лица и в какой-то степени даже испытывать его чувства.

Качество «доброта» характеризует общую установку человека в его отношениях с другими. Однако трудно представить себе возможность «абсолютно доброго человека», отдающего все, всегда, всем и ничего не требующего взамен. Этот идеал, проповедуемый религиями, нереалистичен.

Но что мешает нам сделать абсолютным альтруистом искусственный интеллект? Все во власти конструкторов, можно задать любые чувства и любые их характеристики. Однако как только сознание достигает определенного уровня, появляются самоорганизация и обязательное влияние общества. В этом то все дело. Теперь нам придется разобрать этот вопрос.

Сфера отношений ИИ

Прежде всего отношения предусматривают существование двух или нескольких субъектов, взаимодействующих друг с другом или, точнее, обменивающихся воздействиями — материальными, энергетическими, информационными, смешанными, любыми.

Мы уже рассматривали общий принцип отношений между интеллектом и средой. Интеллект воспринимает воздействия среды, анализирует их, оценивает полезность, планирует и выполняет действия, в результате которых получает «плату», удовлетворяющую потребность и изменяющую чувства. Среда при этом представлялась неопределенной. На рис. 21 среда детализирована — выделена природа, техника, модели и общество. Под последним понимается система, состоящая из интеллектов, подобных данному. Показано, что она дает информацию и получает информацию, следовательно, происходит обмен сигналами. А как же с «платой»? Сигналы оцениваются и могут удовлетворять некоторые потребности (об этом мы уже говорили),

например потребность в новом или потребность в лидерстве. Может статься, что существует и потребность в доброте, потребность отдавать?

До сих пор потребность «отдавать» мы рассматривали в плане действий, которые сопряжены с усилиями, с напряжением. Хотя они сами по себе могут быть приятными, удовлетворяя потребность в деятельности или «рефлекс цели», но при этом не рассматривается объект действий. Его чувства не принимались во внимание. Пока не принимались.

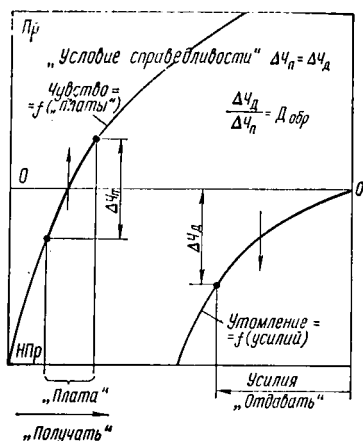
Иное дело — общественная среда, состоящая из людей, а в будущем включающая и искусственные интеллекты.

В описаниях ФА было показано, что при планировании составляется модель объекта, представляющая его состояние после воздействия, — модель-цель. Для материального объекта — это его структура, измененная действием. Речевые ФА предусматривают моделирование реакций человека на сказанную фразу. Цель речевого ФА — вызвать желательные чувства у собеседника и предусмотреть свои собственные чувства в ответ на них («Я ему скажу то-то, он обрадуется, и это будет мне приятно!»).

Самые простые отношения между людьми — это материальный обмен: я даю вам вещь, вы мне взамен даете другую. Предположим, что оба мы делали свои вещи сами, своим трудом. Затраченный труд измеряется утомлением и временем. То и другое измеримо мерой чувств. Мой труд — приращение чувства утомления $\Delta\mathcal{U}_д$, где индекс «д» означает «давать». Вещь, которую я получаю взамен, удовлетворяет какую-то мою потребность и таким образом дает приращение чувства приятного $\Delta\mathcal{U}_п$, где индекс п — «получать». Справедливым обменом я буду считать такой, когда оба приращения чувств по крайней мере равны: $\Delta\mathcal{U}_д = \Delta\mathcal{U}_п$. Втайне каждый хотел бы произвести такой обмен, чтобы $\Delta\mathcal{U}_п > \Delta\mathcal{U}_д$.

Как сравниваются оба приращения чувств? Где мера? Чувства очень разные, и обмены тоже разные. За труд можно получить пищу, вещь или развлечение — все они удовлетворяют различные потребности. Сравнение идет по величине приращения по оси ординат характеристики «плата» — чувство (см. рис. 22), где отложены компоненты интегральных чувств $Pr - NPr$. Оба чувства легко измерить в модели ИИ. А как это делается в мозге?

Рис. 40. Схема, иллюстрирующая «условие справедливости». Показаны характеристики двух чувств: приятного $\mathcal{U}_n$ от получения некоторой «платы» и утомления $\mathcal{U}_d$, связанного с усилиями, которые нужно затратить, чтобы оправдать получение «платы». «Условие справедливости» выполняется при равном приращении чувств — $\Delta\mathcal{U}_n = \Delta\mathcal{U}_d$.



Произведенные этологами наблюдения над животными свидетельствуют о том, что их отношения подчиняются определенным законам. Если одно проявляет благорасположение, другое не бросается в драку, а в худшем случае демонстрирует равнодушие. И наоборот: если одно обнаруживает признаки агрессии, другое отвечает тем же либо убегает. На ласку приблизительно отвечают лаской, на агрессию — сопротивлением. Похоже, что у животных есть мера сравнения чувств от «получения» и «отдавания» и критерий справедливости отношений. Не надо этому удивляться: уже немало чувств, казавшихся сугубо человеческими и привитыми обществом, обнаружено у животных. Следовательно, они имеют биологические корни. Разве мы не видим примеров альтруизма среди стадных животных? Я уж не говорю о проявлениях любви родителей к детенышам — это может наблюдать каждый. На рис. 40 показана схема, иллюстрирующая «условие справедливости». Она построена в предположении, что существует такая следящая система, которая измеряет и сравнивает чувства отношений: приращения приятного от действия «получать» и приращения неприятного от действия «отдавать».

Доброту можно измерить отношением

$$\text{Добр} = \frac{\Delta\mathcal{U}_d}{\Delta\mathcal{U}_n}.$$

У эгоиста существует психологическая установка на отношения, при которых можно давать меньше, чем получать, иными словами выигрывать в чувствах. У альтруиста, напротив, показатель доброты превышает 1. Не следует думать, что альтруизм противоречит основному закону поведения, согласно которому все действия направлены на получение прироста интегрального чувства: $(Pr - HPr) > 0$. Дело в том, что альтруист, затрачивая неприятное чувство, например утомление, на изготовление подарка, получает дополнительно приращение приятного от удовольствия, доставленного другому. Можно сказать, что

$$\Delta\mathcal{U}_n = f_1(\text{«плата»}) + f_2(\Delta\mathcal{U} \text{ другого}).$$

Человек, получивший подарок, выдаст «плату» в словах благодарности (f_1 «плата»), но для альтруиста сопереживание приятным чувствам другого дает еще добавочное приращение чувства приятного (f_2 ($\Delta\mathcal{U}$ другого)). Таким образом, условие приращения суммы приятного в итоге поступка будет соблюдено.

Если сказать предельно просто, то доброму доставляет удовольствие делать приятное другим людям, а эгоист равнодушен к чувствам других и за свои услуги предпочитает получать «плату» «натурой».

К слову, явления сопереживания боли у животных зарегистрированы объективными средствами в опытах физиологов. Они же установили большие различия в силе этого чувства у разных особей. Если продолжить экскурс в этологию, во взаимоотношения животных, то оказывается, что отношения крайне неравномерны. К членам семьи испытываются одни чувства, несколько иные к членам своей стаи, совершенно отличные — к членам другой стаи. Получается, что по отношению к разным особям имеют место совершенно разные «коэффициенты справедливости», разное сопереживание. Чужаков всегда встречают неприязнью. Разве не то же мы видим у людей? Даже в воровской шайке, далекой от всякой морали, ее члены соблюдают своеобразную справедливость по отношению друг к другу. Модельно это выражается разными значениями «коэффициента доброты», разной установкой, что считать справедливым. Видимо, эта разница определяется степенью сопереживания: близким и приятным знакомым мы больше сопереживаем, чем чужим и враждебным нам людям. Играет роль та самая прибавка на чувства партнера по отношениям.

Общеизвестно крайнее разнообразие отношений между людьми, не только чужими, но даже внутри семьи. То же касается и животных — даже сука по-разному относится к своим щенкам. Чем определяются эти различия? Можно предположить несколько факторов:

1. Биологическая близость. Она особенно выражается в близости семейной, в которой играет большую роль инстинкт продолжения рода.

2. Принадлежность к одному рабочему коллективу и к одной социальной группе.

3. Соотношение величины лидерства у вступивших в отношения людей. То же касается и подчиненности.

4. «Набор чувств», структура личности, и в частности — общительность.

5. Общность убеждений.

6. И наконец — обратная связь, самоорганизация.

Пунктов перечислено много, что показывает, сколь сложны отношения между людьми.

Как же быть с искусственным интеллектом? Увы, если он предназначен для человеческого общества, нет другого выхода, как моделировать в нем всю сложность исходных чувств. В частности, необходимо заложить в него «чувство справедливости», должны формироваться симпатии и антипатии, определяющие разницу в «коэффициенте доброты», разницу в значимости сопереживания близким и чужим. Другое дело, что все это при проектировании ИИ проще «дозировать» и можно задавать благоприятные характеристики. Однако и здесь не все однозначно. Осуществить азимовский главный «закон робототехники» — не вредить людям — не так просто. Он упирается в качество воспитуемости: задавать ли возможность изменения «врожденной» характеристики чувств в процессе деятельности или нет, а если да, то в какой степени? Видимо, на этот вопрос нельзя ответить однозначно, потому что возможны разные сферы деятельности ИИ в человеческой среде. Несомненно одно — нельзя допускать излишнюю самоорганизацию искусственного интеллекта, иначе он может выйти из-под контроля людей.

Врожденные потребности-чувства значат для человека много, но не все. Речь и обучение создали возможность приобретать «искусственные чувства», являющиеся выражением потребностей более высокой разумной системы — общества. Как уже говорилось, убеждения — это словесные формулировки, их модели,

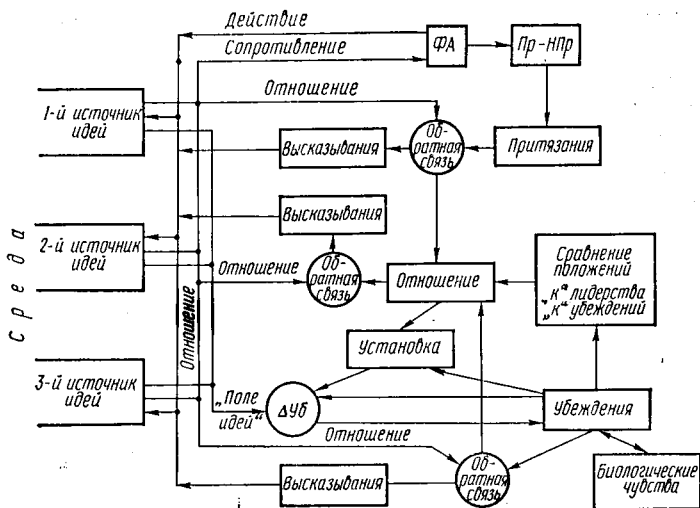


Рис. 41. Схема формирования отношений и убеждений.

обладающие значительной активностью и имеющие сильные связи с центрами интегральных чувств *Пр — НПр*. Они определяют, что можно, что должно, чего нельзя делать, что хорошо и что плохо. Содержание убеждения может быть крайне различным — от запрещения ругаться до отвлеченных идей, побуждающих их приверженцев к самоистязанию или преступлениям против общества.

Формирование убеждений, их изменение в течение жизни в связи с условиями существования трудно формализовать. Тем не менее я сделал такую попытку применительно к некоторому условному интеллекту. Общая схема формирования отношений и убеждений показана на рис. 41.

Социальная среда условно представлена тремя источниками идей. Любых идей, которые выражаются словесными формулами, обращающимися к чувствам. Обычно источниками идей являются окружающие люди — вначале члены семьи, товарищи, учителя, потом — общественные организации, рабочие коллективы, наконец — книги. Но в данном случае не в этом дело, нас интересует схема. Понятие «воспитатель» и «воспитуемый» в равной степени относится и к ИИ.

Каждый из представителей среды, человек или группа людей, имеет свои симпатии и антипатии, свое

отношение к носителю интеллекта. Это отношение он высказывает словами и поступками. Пока ограничим наше рассмотрение словами. Правда, наказания «голодом» и «болью» не исключены, учитывая, что «телесные чувства» будут представлены в ИИ обязательно.

На рис. 41 показаны притязания со стороны интеллекта в адрес своих оппонентов и воспитателей. Понятие это было уже дано. Повторяю: притязания подразумевают величину «платы», которую хотел бы получить интеллект от среды в ответ на свою деятельность с целью максимального удовлетворения своих потребностей. Притязания всегда превышают реальную «плату», потому что если они полностью удовлетворяются, то наступает адаптация чувства и для достижения их максимума нужно увеличивать «плату». Человек живет в условиях компромисса — его притязания всегда превышают «плату», не позволяя адаптироваться к счастью.

Притязания вызывают реакцию на них со стороны адресатов, которая тоже выражается словами и поступками (представьте пример из практики воспитания детей). Первый компонент отношения интеллекта к адресату зависит от этой реакции на его притязания.

Второй компонент — пожалуй, главный — определяется сравнением положения носителя интеллекта с положением соответствующих групп или лиц (снова — адресатами) с поправкой на его лидерство. Отношение высказывается, в ответ на него следует реакция, слушающая обратной связью и воздействующая на отношение.

Так формируется система отношений (симпатия либо антипатия) к разным источникам идей — лицам, социальным группам в зависимости от результата сравнения положений, лидерства и обратной связи.

Отношения не исчерпываются только симпатиями и антипатиями, это еще и авторитеты, и их уровень. Именно они и определяют формирование убеждений.

На рис. 41 все источники идей объединены «полем идей». Оно включает в себя набор словесных формул, их «адрес» (источник) и интенсивность их повторения. Иначе говоря, «поле идей» — это содержание различных убеждений, которые циркулируют в общественной среде, окружающей интеллект. Какие формулы и в какой степени усваиваются — зависит от так называемой установки. Этот термин введен психологами для

обозначения изначального отношения к высказыванию (слушать ли его, воспринимать ли, либо пропустить мимо ушей, как не стоящее внимания или заведомо ложное). Установка зависит от авторитета источника и от степени совпадения содержания идеи с уже имеющимся убеждением, созданным в предшествующие периоды жизни интеллекта.

В памяти развитого интеллекта (а мы говорим именно о таком, когда имеем дело с третьим уровнем сознания) имеется достаточный набор словесных формул, претендующих на роль убеждений,— это и есть все «поле идей» с указанием их адресатов. Некоторые из них уже стали своими убеждениями с определенной значимостью, то есть активностью и связями с чувствами. Внешнее воспитание изменяет эти убеждения на некоторую величину ΔU_6 , зависящую от установки и от интенсивности воздействий (например, от частоты повторения фраз и образов в средствах массовой информации).

Так осуществляется постепенный «дрейф» убеждений: в зависимости от авторитета тех, кто их проповедует, и от содержания уже достигнутых ранее убеждений.

Следует различать содержание убеждений и их значимость, то есть активность по сравнению с врожденными для человека или изначально заданными для ИИ чувствами. Например, можно принимать заповеди религии, но не придавать им особенного значения, и тогда они не могут конкурировать с потребностями, производными от инстинктов.

Убеждения и заданные чувства соединены прямыми и обратными связями: убеждения могут усиливать или тормозить чувства, а те в свою очередь участвуют в формировании убеждений. Проще всего внедрить в сознание те словесные формулы, которые совпадают с насущными потребностями. И наоборот, чем абстрактнее формулы, тем труднее их сделать значимыми убеждениями.

Очень сложно формализовать «поле идей», что необходимо при создании ИИ. В самом деле, существует огромное множество словесных формул, приобретающих значимость чувств, их трудно вместить в модель. Мне кажется, что нужно установить основные направления идей, сопоставить им специальные шкалы, выбрать на каждой шкале противоположные точки и приписать им численные значения со знаками $\pm$

и —. Тогда промежуточную точку можно определять как координату по данному убеждению. Число таких шкал в первом приближении должно соответствовать основным чувствам: жадности, страху, общению, доброте, стремлению к лидерству, любознательности, лени. Люди тоже создают идеи по всем этим чувствам, и общество внушает их своим гражданам, формируя у них убеждения. Не буду разбирать особенности построения шкал по каждому убеждению, это относится к задаче моделирования общественных систем. В качестве примера можно сослаться на «шкалу жадности».

Координаты со знаком «+» характеризуют убеждения, связанные с отказом от вещей, координаты со знаком «—» характеризуют убеждения, оправдывающие собственность.

Следует остановиться на убеждениях, связанных со справедливостью — они, пожалуй, самые важные. Что считать справедливым? Мы уже обсуждали этот вопрос и ввели понятие условия справедливости (см. рис. 40). Обсуждаемые убеждения устанавливают эквиваленты обмена или «ранг ценностей» — труд, власть, вещи, информация. Фактически справедливость сводится к словесным формулам обмена: «что, кому, за что и сколько» давать, «от кого, что, сколько и за что» брать.

Формализацию убеждений для модели ИИ можно осуществить путем введения системы координат с нанесенными на них несколькими вариантами точек. Каждый вариант отражает убеждения, исходящие из определенного источника идей. Собственные убеждения искусственного интеллекта также будут представлены точками на сетке координат.

Убеждения формируются согласно схеме, показанной на рис. 38, как функция воспитания. Однако это происходит лишь в «детстве и юности» искусственного интеллекта, то есть в период его обучения. Приходит время, и уровень сознания ИИ повышается настолько, что он становится способным моделировать отдельно действия, высказывания, чувства и мысли — свои и других лиц (мысли и чувства других — по косвенным признакам их речи и мимики). Тогда наступает время самоорганизации или применительно к убеждениям — время самовоспитания. Убеждения (точки в упомянутой системе координат), высказываемые различными адресатами, сравниваются с их действиями и вводится поправка на справедливость (искренность),

Образование и накопление знаний изменяют баланс состояний, и авторитеты падают. Устанавливается уровень «биологических», то есть изначально заложенных чувств, и они влияют на пошатнувшиеся убеждения. Накапливается собственный опыт действий как по «убеждениям», так и по врожденным потребностям, отмечается уровень душевного комфорта, сопровождавший эти действия. Все это вносит коррективы и смещает точки в системе координат, чаще в сторону изначально заложенных чувств. Происходит «эрозия убеждений». Но может произойти и обратный процесс: действия по убеждениям усиливают их модели, они становятся более значимыми и меняют «ранг ценностей» в сторону идеалов.

Знания, чувства, убеждения, характер и отношения к окружающим искусственного интеллекта делают его индивидуальностью, личностью, причем личностью социальной. Высокий уровень сознания выводит его за пределы объекта простых манипуляций со стороны воспитателей. И здесь возникает ряд вопросов.

Нужны ли убеждения для искусственных интеллектов? Они нужны, во всяком случае для той их разновидности, которую мы обсуждаем, то есть для предназначенных для деятельности среди людей и в помощь людям. В речевом поведении в значительной мере они заменяют чувства.

Нужны ли воспитуемость, самоорганизация и тем более — самовоспитание? Этот вопрос гораздо сложнее. «Законы робототехники» очень логичны, хотя и трудно выполнимы при высоком уровне сознания и большом объеме знаний. Формирование личности искусственного интеллекта будет проходить ускоренными темпами, так же как и обучение. Следовательно, можно довольно жестко задать характеристики изначально заложенных потребностей и в короткий срок создать у него желательные для людей убеждения. Последующая возможность их изменений в результате анализа действительности и собственной деятельности, к сожалению, не может быть заблокирована. Однако жесткие основные чувства, такие, как альтруизм, общительность, должны обезопасить людей от чрезмерной увлеченности искусственного интеллекта своими вновь изобретенными идеями.

Еще одной гарантией безопасности должны быть ограничения тренированности моделей или их максимальной активности. Это средство против излишней

увлекаемости и в то же время условие психической нормальности ИИ. О возможности психических заболеваний уже была речь.

Как упоминалось, искусственный интеллект такого уровня сознания и образования представляет собой личность. Людям придется с этим считаться и рассматривать его как равного. Это будет существо, чувствующее так же, как и человек, только еще умнее, поскольку он будет обладать гораздо более совершенной памятью. Поведение такого существа можно представить себе как выражение общей программы всех интеллектов: выбор и смена деятельности с целью получения максимума УДК в каждый данный момент. Это вовсе не значит, что счастье данного момента будет законом его деятельности. Как умный человек планирует длительные ФА и живет будущими их результатами, а в настоящем напрягается и ограничивает себя, так же будет жить и искусственный интеллект.

Чем он будет заниматься? Тем же, чем и люди: по «расписанию» внешней среды выполнять свои социальные обязанности, то есть работать, а в остальное время — развлекаться, учиться, общаться с людьми или другими искусственными интеллектами.

Еще один вопрос: можно ли будет корректировать искусственный интеллект в процессе его деятельности для людей, если она окажется сомнительной? Например, изменить активность тех или иных чувств? Можно, но только при одном условии — его собственном согласии. Здесь мы сталкиваемся с моральными и этическими проблемами. Искусственный интеллект третьего уровня сознания придется рассматривать как существо со всеми правами человека. Нервнобольных лечат, но только тогда, когда они этого хотят сами. Психически больных лечат без их согласия, по просьбе родственников, врачей, общественных организаций. Точно так же должно обстоять дело и с искусственным интеллектом. Его право на жизнь будет таким же, как у людей. «Выключить» его можно будет только по приговору суда... Странные проблемы возникнут! Боюсь, что серьезный читатель обвинит меня в излишнем воображении и соскальзывании с научных позиций на банальную фантастику. Это не так. Все, о чем я пишу, почти так же реально, как использование термоядерной энергии в мирных целях: никто не сомневается, что это возможно, а дело пока стоит, упирается в технологические трудности. Я уверен, что алгоритм ин-

теллекта вполне возможен в самых высших его проявлениях и что создание технологии для этого тоже возможно. К сожалению, одной уверенности здесь недостаточно, пока не вырисуются перспективы практической реализации такого алгоритма.

Общество искусственных интеллектов

Никак не обойтись без этой главы, выглядящей еще более фантастично и несолидно для научной книги. Речь пойдет об обществе (или поскромнее — коллективе) ИИ. Не нужно удивляться: это перспектива не безнадежно далекая — в пределах жизни молодых (40—50 лет).

Первая возможность — взаимодействие ИИ, созданных для помощи людям. Вторая — колонии роботов-интеллектов на планетах. Наверное, это будет проще, чем заселять их людьми, — очень уж хрупка наша живая плоть.

Зададим сначала вопрос: в чем будет «смысл жизни» искусственных интеллектов?

А в чем смысл жизни людей?

Нет, я отнюдь не собираюсь философствовать. Модельное представление об интеллекте ставит вопрос совсем в другой плоскости. Жизнь отдельного человека — это деятельность, поведение, поступки, в том числе и речевые. Это иерархия и сеть ФА, направленных на удовлетворение биологических потребностей, измененных общественным воспитанием, социальных потребностей — убеждений, целиком привитых обществом. Такая деятельность имеет эгоистическую направленность, в лучшем случае — на собственное потомство и непосредственное окружение. Деятельность индивидуума, личности, живущей по убеждениям, является проявлением жизни общества, так же как деятельность клетки хотя и самостоятельна, но целиком зависит от целого организма и служит его функциям. Чем больше прогрессирует человечество, тем дальше в будущее нацеливается жизнь общества, тем шире становится круг связей государств и людей. Таким образом, и деятельность каждого человека — его ФА — все более утрачивает эгоистическую направленность, и смысл его жизни (хотя и не осознаваемый) смещается к служению обществу, как высшей системе.

Относительно недавно для большинства людей главным объектом приложения их деятельности была природа. Жизнь состояла из непосредственного добывания пищи и защиты от врагов — холода, зверей. Правда, люди уже давно объединились для борьбы с природой, но общение между ними играло подсобную роль. Теперь все изменилось совершенно. Главным объектом деятельности большинства людей в высокоразвитых странах являются сами люди и созданная ими техника. Непосредственно с природой общается только несколько процентов населения, но и те — чаще при помощи техники, чем прибегая к собственной мышечной силе. Содержанием жизни большинства стали общественные отношения. Если самый глубокий смысл жизни — в его модельном выражении — остался таким же: стремление к максимуму уровня душевного комфорта, то компоненты, его составляющие, стали иными. Подавляющая часть потребностей теперь удовлетворяется через отношения людей. Мы даже мыслим больше словами, чем образами.

Все это я говорю для того, чтобы подойти к смыслу жизни искусственных интеллектов как членов общества. Пока они будут вкраплены в человеческое общество, они будут жить смыслом человеческой жизни. Представьте себе общество бесполой людей — это совсем не так трудно вообразить. Людям было бы достаточно занятий, развлечений и удовольствий и за пределами инстинкта размножения. По условиям возрастным и социальным многие живут такой «бесполой» жизнью. Так же, теми же стимулами будут жить среди нас искусственные интеллекты. Будут ли они счастливее нас, людей, зависит от того, как их спроектировать. Уже говорилось о том, что есть некоторые непреложные законы деятельности интеллекта, обязательный набор потребностей, свойств, в том числе скука и адаптация, без которых ИИ невозможен, но которые не позволяют ему быть безоблачно счастливым.

Предположим, поставлена задача создать роботов с интеллектом, которыми нужно заселить планету, причем так, чтобы колония могла жить самостоятельно и развиваться без связи с людьми и Землей. Основное требование одно — прогресс, возрастание и развитие общества в целом. Другое требование (сомнительное и спорное) — члены общества должны быть по возможности счастливы: с некоторым достаточным

УДК, со сравнительно небольшим его разбросом среди групп «населения».

Как для таких ИИ спроектировать интеллект — вопрос не главный. Он может действовать только по одним законам, если говорить о развитом интеллекте, с третьим уровнем сознания. Эти законы были обсуждены выше. Вопрос — в потребностях и чувствах, которые будут двигать интеллектом, в качествах, которые нужно ему придать, чтобы общество удовлетворяло требованию прогресса и устойчивости.

Вернемся к смыслу жизни — уже для членов колонии ИИ. Борьба с природой? Да, борьба останется. Нужны энергия и материалы для того, чтобы воспроизводить самих роботов, огромную техносферу, обеспечивающую получение этих материалов и энергии. Прогресс общества — в науке, а она материальна. Уж кто-кто, а ученые знают, что наука может поглотить сколь угодно техники и энергии. Следовательно, будет большая и очень разносторонняя «материальная» в буквальном смысле слова деятельность. Ей обязательно сопутствует или даже предваряет ее информационная индустрия, тоже требующая материальных средств.

Другие точки приложения сил для ИИ будут примерно такими же, как и у нас. Общение. Искусство воспитания новых поколений. Поглощение информации. Соппротивление старению и болезням. Впрочем, последняя проблема будет выглядеть иначе, чем у людей, хотя и не исчезнет совсем. Вот и все аспекты смысла жизни.

Все это возможно только в очень развитом технологическом обществе. Сложность его, если включить науку, будет очень велика. Можно ли рассчитывать, что эта сложность влезет в интеллект каждого члена колонии роботов? Думаю, что нельзя. Внешняя память — библиотеки информации и программ — будет доступна всем, а вот собственная память не может содержать всех моделей. Общество в целом должно быть гораздо умнее своих членов. Еще одно условие, вытекающее из разнообразия и сложности, — технологическое неравенство членов общества. Одни должны будут выполнять одну работу, другие — другую. Если невозможно вместить все модели в интеллект каждого, то потребуются «специалисты».

Другой вытекающий из этого вопрос: может быть, следует проектировать разных роботов — «элиту», 190

«рабов»... Не думаю. «Рабами» должны быть машины с элементарным интеллектом без всяких признаков самоорганизации, но, возможно, с некоторым самым низким уровнем сознания. Такое же положение, как теперь с рабочим скотом. Машины будут даже еще гораздо ниже — просто умные автоматы. (Уверен, что уровень сознания высших млекопитающих, к примеру, собак, лошадей, живущих с людьми, не так уж низок и мы повседневно нарушаем их «права»).

Но как только создадут ИИ, способный к обучению, уже невозможно будет удержать его на уровне «раба». Поэтому самое разумное для удовлетворения второго требования (обеспечить достаточный УДК членов колонии) — это «врожденное равенство» интеллекта. Может быть, полная идентичность? Не думаю. Нужно разнообразие, обеспечивающее различия в интеллектуальной деятельности, но не более. Нельзя примирить большое неравенство, способность к сопереживанию, высокий уровень сознания и высокий УДК!

Попытаемся представить себе необходимый набор потребностей-чувств, который будет достаточен, чтобы обеспечить прогресс, высокий УДК членов колонии и устойчивость системы. Одно условие уже высказано — набор чувств одинаков у всех, а значимость их различается у индивидов в некоторых допустимых пределах. Эти пределы еще требуют проработки на моделях.

Потребность самосохранения несомненно останется, как уже доказывалось применительно к искусственным интеллектам, назначение которых — служба людям. Нужен «страх» и нужен «голод», так как возможна трудная жизнь, с опасностями. Нужны эмоции с некоторыми ограничениями пределов.

Рабочие потребности уже несколько раз перечислялись. Вот они еще раз: любознательность и потребность поиска, потребность действовать и доводить ФА до конца — «рефлекс цели», «рефлекс свободы». Потребность в отдыхе и расслаблении как реакция на напряжение и однообразие.

Интегральные чувства *Пр* — *НПр* обязательны. Именно по ним интеллект оптимизирует свою деятельность.

Такие качества, как адаптация и способность моделей и связей к тренировке, предопределяющие самоорганизацию, являются неременной принадлежностью всякого развитого интеллекта.

Самый спорный вопрос — это общественные чувства и потребности. Как было бы хорошо представить себе благополучное общество, в котором «граждане» с искусственным интеллектом лишены всех пороков, портящих людям жизнь. Однако будет ли состоящее из таких «граждан» общество удовлетворять требованиям прогресса, «счастья» и устойчивости — в условиях исключительной сложности среды и структуры самого общества?

Рассмотрим вопрос о лидерстве и подчиненности. Несомненно, деятельность будет коллективной, поскольку сложность среды во много раз превысит моделирующие возможности каждого из ИИ. Следовательно, моделирование будет осуществляться с упрощениями, а значит, и с искажениями, вызванными субъективностью. Субъективность, ограниченность ведут к разногласиям в оценках. При этих условиях эффективная деятельность коллектива возможна только при наличии руководителя, лидера. Назначение его из числа равных не решает проблемы: нужен авторитет, только тогда коллектив работоспособен. Как ни рассуждай, но без потребности руководить и подчиняться, выраженной в разной степени у членов любого сообщества, работоспособный коллектив невозможен. Итак, эти потребности необходимы. Лидерство должно сочетаться с повышенными способностями, подчиненность — со способностями поскромнее. Тогда все станет на свои места.

Однако — при одном условии: лидерство должно быть ограничено. Нужно задать такие характеристики и так лимитировать тренируемость центров, то есть способность изменяться в результате самовоспитания, чтобы не развился диктатор.

Есть в арсенале чувств противоядие против излишнего лидерства — это сопереживание, альтруизм (доброта) и потребность в общении, чувство справедливости и правды. Вместе взятые, они дают достаточную основу для правильной морали, которая должна детализироваться убеждениями, привитыми воспитанием. Если при проектировании ИИ соблюсти баланс между потребностями, способными перерасти в пороки, разъединяющие «граждан», и потребностями, их объединяющими, то можно приблизиться к идеалу. Приблизиться, но не достигнуть полной гармонии, потому что тогда наступит застой. Видимо, нужно предусмотреть достаточное разнообразие интеллектов. Конечно, это

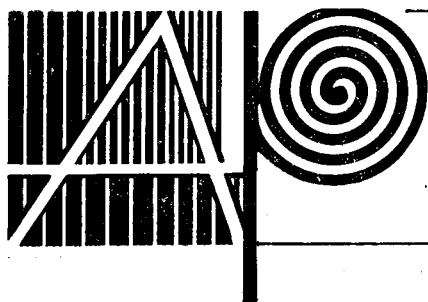
породит противоречия, но без противоречий нет прогресса.

Интересны проблемы воспитания, образования и последующего самовоспитания ИИ. Несомненно, нельзя сразу изначально задать все чувства, убеждения, память, воспоминания, которые определяют личность. В то же время нелогично заставлять «искусственного ребенка» проделывать самостоятельно и медленно весь цикл естественного развития в обществе, подвергаясь случайностям и опасностям неправильного направления самоорганизации. Безусловно, многое нужно задать в первичной структуре — в связях с высокой проходимостью и в характеристиках элементов, например в виде «центров» потребностей и чувств. Это врожденные, «генетические» программы. Последующее развитие следует организовывать в виде обучения и общения, в процессе которого в соответствии с условиями общества должна формироваться личность с особенностями, определяемыми ее врожденными качествами. Ясности в этом вопросе нет. Можно ли на этот период задать специальные условия образования связей, обеспечивающие быстрое обучение и формирование убеждений? Хотя на первый взгляд кажется, что все в нашей власти, но это не совсем так. Трудно представить, например, оперативную память с необозримым объемом или возможность изменения свойств элементов — все это ограничено технологией. У животных и человека усвояемость в детстве выше, чем у взрослого, но еще не ясно, как это осуществить для ИИ. Одно несомненно, что образование и воспитание «искусственных детей» можно запрограммировать гораздо жестче, чем для человеческих, тем более, что они не будут ограничены рамками роста «тела».

Жизнь человека и даже исторические периоды человечества очень коротки по сравнению с эволюцией природы. И ничего пока сделать не удастся для ее ускорения. «Генная инженерия» — не утопия, но возникающие здесь проблемы гораздо сложнее, чем представляем мы это сегодня. Для целенаправленной переделки гено типа человека нужна его кибернетическая «действующая модель» огромной сложности. Это значительно труднее, чем создать искусственный интеллект. Опыт развития науки и техники показал, что люди не повторяют природу, а изобретают новые решения. Уверен, что проще сделать «нового человека», чем переделывать имеющегося.

Не приведет ли высокий уровень технологии к искушению быстрой переделке природы «граждан» искусственного общества? Не обусловит ли это его нестабильность? Разумеется, сейчас не стоит серьезно обсуждать этот вопрос, но такая опасность возможна. Консерватизм необходим, потому что явления, происходящие в системах «типа живых» — в организме и обществе, слишком сложны и их отражение в разуме типа человеческого весьма ограничено и субъективно. Однако уже оформляется модельный подход к сложности, который позволяет исправить недостатки разума. Нет сомнения, что будущее общество предусмотрит опасности неконтролируемой самоорганизации.

Фантастична и увлекательна проблема искусственного интеллекта, поэтому меня и занесло так далеко — до разговора об искусственном обществе. Вернемся к вопросам теории интеллекта.



На пути к интеллекту выше человеческого

Разум человека и сложные системы

Мы уже разобрали два уровня интеллекта — уровень животного и уровень человека. Намечали пути к созданию алгоритма, способного обеспечить их моделирование. Напомню специфику интеллекта человека. Он управляет мышцами и через них — средой по критериям, исходящим от тела, от самого разума, от семьи и от общества. По своему алгоритму разум был «рассчитан» природой так, чтобы обеспечить оптимальную деятельность в изменчивой внешней среде умеренной сложности через короткие ФА. Этим определились его недостатки, ограничившие его же объективную ценность. Первый недостаток — меняющаяся субъективность оценок, связанная с переменной активностью критериев-чувств. Второй недостаток — неоднородность и неравномерность моделей среды: область центрального зрения представлена подробно, периферия — неясно. Проблема центра и фона объясняется рядом причин: спецификой зрительного восприятия и механизмом сознания, выделяющего одну модель в каждый момент, а также изменчивостью чувств, усиливающих одни модели в ущерб другим.

Модели в коре мозга можно рассматривать как действующие, поскольку они воспроизводят динамику среды, но отражают ее крайне неравномерно: для центральной фигуры улавливаются подробности движения, периферия представлена обобщенно и фрагментарно.

Кроме субъективности и ограниченности моделирования, мозгу присуща еще и увлекаемость, то есть способность к повышенной тренировке некоторых моделей в результате их повторного привлечения в сознание. При этом они сами становятся фактором, направляющим движение активности по моделям. Такой процесс и является основой самоорганизации. Она полезна для приспособления, но только в некоторых пределах, и сама может стать источником искажения действительности.

Наконец, речь как искусственный код моделей, удобный для их передачи вовне и хранения, далеко не полностью соответствует образам. Во-первых, речь отражает только сознание, во-вторых, она «беднее» образов, в-третьих, статична и медленна. И уж конечно, очень субъективна: при передаче моделей от одного лица к другому истина заметно искажается. Отсюда следует недостаточность словесной информации для описания сложных систем. По ней нельзя построить действующие модели большой сложности. Математика и графика частично исправляют эти недостатки речи, но лишь частично, поскольку все внешние модели действуют только через разум, а его недостатки — субъективность, увлекаемость и ограниченность — непреодолимы.

Длительная память человека несколько исправила недостатки «алгоритма» разума, но тоже не полностью. В частности, он оказался малопригоден для моделирования сложных систем в целях науки.

Специфику сложных систем «типа живых» я уже описывал, но, наверное, стоит повторить основные пункты. Эти системы сложны, так как состоят из множества элементов, по-разному комбинирующихся друг с другом в различных количествах. Они имеют рабочие подсистемы, в которых осуществляются главные энергетические и вещественные взаимоотношения с внешней средой, и управляющие, в которых представлены модели, взаимодействующие друг с другом сигналами. Все сложные системы построены по иерархическому принципу, причем на каждом уровне есть свои коды моделей и сигналов. Подсистемы управления имеют еще собственную иерархию моделей. «Вертикальные» и «горизонтальные», прямые и обратные связи объединяют систему в одно целое. Прочность этих связей различна, чем определяется степень системности целого или, иными словами, степень авто-

номности частей, их «жизнеспособности» вне целого. Наконец, существует самоорганизация как способность изменять свои функции и структуры под влиянием внешней среды и собственной деятельности для удовлетворения главного критерия системы — выживаемости. Можно было бы продолжить перечень качеств сложных систем, но и эти я повторил лишь для того, чтобы перейти к вопросу о требованиях, предъявляемых к интеллекту, нужному для их познания.

Как уже говорилось, познание — это моделирование, создание моделей. Для этого недостаточно просто восприятия, нужны исследования, то есть активные воздействия на объект по плану, составленному на основе гипотезы. Тут познание смыкается с управлением. Интеллект, работающий по типу животного или даже человеческого разума, мало приспособлен как для познания, так и для управления сложными системами, поскольку он не в состоянии объективно и точно отразить их сложность в своих моделях.

Выше говорилось, что для эффективного управления объектом нужны его разные модели: обобщенные — для выбора «стратегии», то есть для планирования ФА, охватывающих длительные процессы, частные — для управления в ограниченных пределах и полные (с разной степенью полноты) — для создания материальной копии и для улучшения системы по избранным критериям.

Разум человека не приспособлен для создания таких полных моделей. Более или менее простые объекты удалось смоделировать, благодаря совершенствованию «внешних» моделей — математических, графических и описательных. Однако все они статичны и «для оживления» должны восприниматься разумом. Кроме того, графики, формулы и слова пригодны лишь для представления простых моделей, потому что при усложнении они быстро становятся необозримыми. Именно по этой причине до сих пор ни одна теория сложных систем «типа живых» не доведена до такой полноты и строгости, как теория физических явлений. Можно предположить, что при существующих возможностях человеческого разума и средствах внешнего моделирования (речь, математика, графика) есть «барьер познаваемости» сложности. Я не рискую категорически утверждать, что он непреодолим, но кажется, что это так. Коллективное мышление здесь не поможет. Разделив сложную систему на множество частей

и сделав каждую из них порознь объектом специального изучения, можно уменьшить количественное несоответствие между уровнями сложности объекта и разума. Но это не решает проблемы, потому что в сложных системах поведение части зависит от целого вследствие наличия прямых и обратных «вертикальных» связей между иерархическими уровнями структуры. Следовательно, для познания такой системы нужен синтез.

Вывод отсюда только один: нужны действующие модели большой сложности, то есть принципиально новый код внешних моделей. Математика предлагает аппарат для такого кода, но она статична, и хотя формулы могут отразить динамику объекта, но ручные расчеты по ним весьма ограничивают пределы сложности.

Электроника и ее воплощение в компьютерах предоставляют нам принципиально новые средства создания сложных действующих моделей. Во-первых, они усиливают традиционную математику, поскольку ускоряют счет. Во-вторых, они дают новые возможности для коллективного творчества в виде объединения моделей в единое действующее целое, которое, в пределе, может стать недоступным для полного понимания каждым из участников создания моделей. И наконец, появилась возможность самоорганизации и саморасширения моделей, когда они из средств выражения человеческого разума сами приобретают его качества — способность создания новых моделей.

Действующие модели сложных систем — новый и высший этап оптимального управления этими системами. До тех пор, пока их модели представлены в книгах и «оживают» только в мозге человека, управление остается ограниченным и будет сопровождаться ошибками.

К сожалению, построение таких моделей сопряжено с исключительными трудностями. Дело в том, что они должны быть количественными, как это имеет место в реальных объектах. Существующие ныне науки о сложных системах «типа живых» являются в основном описательными. Они даже не имеют непротиворечивых гипотез по самым основным вопросам функций, не говоря уже о крайней скудости количественных сведений. Первое связано со вторым: для доказательства гипотезы нужно много сопряженных количественных данных, а для их сопоставления, со-

пряжения необходима модель, то есть обобщенная гипотеза.

Становление любой науки можно представить в такой последовательности:

а) наблюдения объекта органами чувств и словесное описание его структуры и функции;

б) измерения функций и уточнение структур — сбор разрозненной количественной информации;

в) синтез «полной» модели объекта, представляющей его теорию с различной степенью обобщенности. Я называю эту модель реальной.

Все науки о живых системах находятся приблизительно между второй и третьей фазами: сведений уже много, а полные модели только начинают создаваться. Процесс этот оказался очень трудным. По существу, еще нет ни одной полной модели. Мне кажется, что нужен новый подход к этой проблеме. Мы предлагаем его в виде метода эвристического моделирования.

Метод эвристического моделирования

Принцип метода состоит в том, что создается математическая модель объекта на основании описательной гипотезы о его структуре и функциях с использованием имеющихся в литературе количественных данных и добавлением недостающих путем предположений, построенных исходя из гипотезы и качественных сведений. Естественно спросить: зачем нужна такая модель и чем она лучше словесного описания? Конечно, она не является реальной моделью. Однако создание ее имеет смысл и представляется мне неизбежным этапом на пути построения реальной модели. Значение эвристической модели в следующем:

а) она требует формулирования более или менее непротиворечивой гипотезы. Противоречия неизбежно вскрываются, когда при построении модели слова приходится заменять цифрами, а также при дальнейшем исследовании готовой модели. Важно, чтобы она вела себя адекватно объекту по возможности в широком диапазоне режимов;

б) создается язык будущей реальной модели;

в) модель четко формулирует задачи для экспериментов: нужно получить определенную количествен-

ную информацию для уточнения наиболее спорных мест;

г) по мере получения новых экспериментальных данных гипотетическая модель приближается к реальной;

д) модель можно исследовать вместо объекта, и она позволяет предположить его новые свойства;

е) наконец, ее можно использовать для управления объектом в тех пределах ее деятельности, где она достаточно точно совпадает с объектом.

Конечно, значимость отдельных пунктов меняется в зависимости от объекта.

Для создания эвристической модели предлагается типовой план:

1. Формирование цели работы или назначения модели: например, как этап в изучении объекта, как инструмент управления, для отработки языка, для проектирования экспериментов и пр. От цели зависит все последующее.

2. Выбор уровня модели. Все сложные системы построены по иерархическому принципу. Степень обобщенности модели определяется тем нижним структурным уровнем, начиная с которого модель должна воспроизводить объект. Уровень определяется назначением модели, наличной информацией и возможностями ее переработки. Для управления достаточны высокие уровни, для создания новой системы и ее изучения желательны, по возможности, низкие уровни. Пример: для понимания механизмов рака нужно моделировать организм с уровня макромолекул, а для управления кровообращением достаточно начинать с уровня органов.

3. Формирование качественной гипотезы о структуре и функциях объекта в пределах, ограниченных целями. Обычно приходится выбирать между несколькими противоречащими друг другу гипотезами. Первый выбор определяется общей точкой зрения авторов. В последующей работе гипотеза подвергается изменениям, если возникают непримиримые противоречия.

4. Построение блок-схемы объекта. Элементы, подсистемы и связи определяются гипотезой и выбранным нижним уровнем структур.

5. Выбор значимых переменных (ограничение числа связей). Сначала определяются все известные переменные для каждого из элементов, потом выбираются, 200

согласно гипотезе, значимые с учетом поставленной задачи. Таким образом, уточняются связи и строится структурная схема объекта, которая становится основной модели.

6. Установление по тем же принципам внешних «входов» системы — сначала определяются все внешние воздействия, потом из них выбираются значимые для сформулированных целей. Устанавливаются граничные условия.

7. Установление характеристик элементов, то есть зависимостей «входы» — «выходы» и «время». Это наиболее произвольный и сложный этап работы, так как данные литературы либо противоречивы, либо недостаточны, либо вообще отсутствуют. Статические и динамические характеристики каждого элемента могут быть выражены графиками, алгебраическими или дифференциальными уравнениями, их системами.

8. Отладка модели. Задаются начальные внешние условия, исходное состояние элементов и производится «увязка» всех характеристик. При этом производится согласование «входов» и «выходов» как целой системы, так и ее элементов. В ходе такой работы обнаруживается противоречивость характеристик некоторых элементов при крайних режимах, требующая коррекций. Иногда возникает и полная невозможность сбалансировать модель, указывающая на непригодность принятой гипотезы. Отладка производится для нескольких граничных условий. Для сложной системы «типа живых» принципиально невозможно создать идеальную модель, так как нельзя повторить все ее низшие уровни.

9. Исследование модели, то есть просчитывание многочисленных статических и динамических режимов, что осуществимо только при использовании вычислительных машин. Сначала нужно создать и отладить программу, что обычно требует некоторых коррекций в самой модели, прежде всего исправления характеристик элементов (например, приведения их к линейным). Само исследование уже позволяет получить новую информацию об объекте, предположить неизвестные до того качества.

10. Верификация модели — сравнение характеристики модели и объекта при одинаковых условиях, с целью определения достоверности модели и особенно

Эвристические модели некоторых сложных систем

В отделе биокibernетики Института кибернетики АН УССР за последние годы была проделана большая работа по созданию эвристических моделей. Описание каждой из них потребовало бы специальной главы; сведения о них опубликованы, поэтому я ограничусь лишь перечислением. Наиболее значительными были модели искусственного интеллекта, но об этом уже шла речь.

Следующей работой является «Модель личности», призванная воспроизвести поведение человека в самом обобщенном виде, что необходимо для создания моделей социальных систем. «Выходы» модели состоят из двух разделов:

а) напряжение и продолжительность основных видов деятельности или распределение труда и времени — на работу, дом, общение, развлечения, информацию, отдых;

б) высказывания и поступки «за» и «против» по отношению к семье, коллективу, обществу, своей и другим социальным группам.

«Входы» представлены воздействиями общества в виде различного рода шкал «платы» со знаком «+» или «—», направленными на удовлетворение основных биологических и социальных потребностей личности. Стимулами деятельности являются чувства — производные удовлетворения потребностей, которые в свою очередь представляют собой функцию интенсивности труда и шкалы «платы» за него. В зависимости от врожденной и воспитанной активности потребностей-чувств, шкал «платы» и труда частные чувства изменяются в пределах некоторых величин от *НПр* до *Пр*. Сумма чувств составляет вместе уровень душевного комфорта, который и является главным показателем субъективного состояния. Таким образом, модель личности связывает общество с его моральными и материальными шкалами «платы» за деятельность, с психикой человека и его трудом. В свою очередь сами шкалы являются функцией труда граждан, уровня развития техники и совершенства общественного устройства, основанного на определенной идеологии.

Модели обобщенных личностей социальных групп могут служить базой для построения «Модели общества». Эта работа начинается в отделе биокибернетики Института кибернетики АН УССР.

Нами использован принцип эвристического моделирования и в сфере физиологии, при создании «Модели внутренней сферы» организма, воспроизводящей взаимоотношение органов между собой, с нервными и эндокринными регуляторами. Практически это означало моделирование кровообращения, дыхания, водного, солевого и энергетического обмена и терморегуляции. В первую очередь модель отработывала норму — физические нагрузки разной интенсивности при неодинаковой внешней температуре. Эта работа проделана, и получены удовлетворительные совпадения с опытом. Главной целью остается моделирование патологических состояний с тем, чтобы в практике использовать модель для управления лечением больных с острыми расстройствами в результате травм, операций, инфарктов и пр. Такая модель представит высший этап медицинской кибернетики, воспроизводящей не статистику болезней, как делается до сих пор, а физиологические механизмы нормальных и патологических процессов. Она может претендовать на определенный уровень теории физиологии. В качестве низшего уровня структур приняты органы, а клеточные механизмы учтены в обобщенном виде в характеристиках органов. Конечно, такая модель пригодна только для описания органной физиологии и не может объяснить, например, механизмы рака или склероза.

Я не буду подробно обсуждать математические проблемы эвристических моделей и ограничусь лишь кратким перечислением условий, связанных с их спецификой.

1. Много переменных. Количество их определяется назначением модели и наличием данных. Так, для физиологических моделей, больше других претендующих на приближение к реальным, количество переменных составляет несколько сотен, поскольку для дальнейшего увеличения их числа просто нет достоверной информации (например, чтобы «спуститься» с уровня органов на молекулярный). Модели интеллекта не рассчитаны на воспроизведение процессов в мозге, но количество «слов», которыми необходимо манипулировать доказательно, чтобы смоделировать мышление человека, видимо, должно исчисляться

многими тысячами. Напротив, модель личности можно ограничить сотнями переменных, так как она по своему назначению предполагает высокую обобщенность и связана с ограниченными возможностями лабораторной оценки психики. Другое дело — общественные системы. Их объем, видимо, должен быть весьма велик.

2. Сложные системы иерархичны по своей структуре, содержат множество «горизонтальных» связей в пределах уровня и «вертикальных» — между ними. Переменные на разных уровнях имеют разную специфику и временные характеристики. Все это должно быть представлено в модели, иногда при помощи и дополнительных переменных, отражающих качество основных.

3. Как правило, характеристики элементов нелинейны. Степени их нелинейности крайне различны, и некоторые точки кривых целесообразно выражать «скачками» (или логическими переключениями), отражающими дискретность в деятельности систем. Примером является переключение СУТ.

4. Необходимость обобщать переменные, то есть заменять несколько конкретных переменных одной обобщенной (условной), является неизбежной при моделировании. Нужны специальные правила, описывающие, что можно, а что нельзя объединять. По всей вероятности, они должны основываться на корреляциях показателей.

5. В эвристических моделях нет необходимости в точности вычислений, поскольку ее нет в экспериментальных науках, изучающих моделируемые объекты. Это очень важное условие, так как оно позволяет отказаться от сложных математических описаний. Так, например, можно отказаться в ряде случаев от дифференциальных уравнений в пользу алгебраических и динамику систем рассчитывать по временным тактам. Нелинейные характеристики можно заменять кусочно-линейными.

6. Модели должны предусматривать вероятностные расчеты. Поскольку в системах-объектах очень много неизвестного, то неизбежно несколько вариантов допущений, существенно влияющих на поведение системы. Так, например, в модели внутренней сферы, призванной воспроизводить динамику развития болезни, подобные варианты совершенно необходимы. То же касается моделей общества. Иное дело — искусствен-

ный интеллект, который можно создать строго детерминированным.

7. Специфика метода эвристического моделирования предъявляет свои условия к программированию моделей на ЦВМ. Программы должны быть гибкими, блочными, позволять произвольно изменять любую величину, любую характеристику. Это необходимо для процесса создания самой модели. Задача разработчика программы не ограничивается воспроизведением заданных формул и цифр, часто приходится их заново создавать и вносить поправки в ходе отладки модели, с тем чтобы получить некоторые предполагаемые по гипотезе конечные «выходы».

Создание эвристических моделей — творческая работа коллектива специалистов в данной области науки и математиков. Те и другие должны проникнуться общими идеями и достигнуть полного взаимопонимания. Роль ведущего в группе определяется не специальностью, а способностью широко охватить предмет и создавать гипотезы. Конечно, нужны также работники-эрудиты, хорошо ориентирующиеся в массе имеющихся фактических данных, программисты, кропотливо отлаживающие сложные программы и готовые в любой момент переделывать их заново в связи с изменением гипотезы.

Эвристические модели приближают нас к теории систем «типа живых», позволяя прогнозировать их поведение, исследовать возможности управления и даже реконструкции. Более того, эвристические модели обещают совершенно новый аппарат познания. Такие модели систем «типа живых» составляют основу для построения в будущем реальных моделей, призванных заменить традиционные книжные модели нашей науки. Разработка эвристических моделей интересна сама по себе, поскольку удовлетворяет чувство любознательности. В самом деле, что может быть заманчивее, чем попытаться заглянуть в механизм работы клетки, целого организма или понаблюдать поведение человека с заданными генами?

Разумеется, реальные модели систем «типа живых» такой сложности, чтобы по ним можно было создавать новые объекты и даже реконструировать их, — дело далекого будущего. Однако искусственный интеллект выше человеческого разума отстоит во времени, пожалуй, еще дальше.

Мне представляется, что для сложных объектов будет целая система действующих моделей — полных, разной степени обобщенности, и частных, в которых воспроизведены детали. Модели эти отразят разные уровни структурной иерархии. Например, можно представить себе действующую модель организма как целого — с его «входами» извне и «выходами» в виде поступков. Наша обобщенная модель личности примерно соответствует этому понятию. Мыслима действующая модель организма на уровне органов — это наша модель внутренней сферы в самом первом приближении. Конечно, в биологии главной должна быть действующая модель клетки как самого низкого структурного уровня, на котором и осуществляются все биологические процессы. Они еще недоступны.

Как бы ни были сложны модели, они никогда не могут стать копией живой клетки или организма, даже если для копирования будут использованы гены данного живого существа. Поэтому модели будут всегда лишь вероятностными. Для того чтобы использовать такие модели в целях управления, придется их «привязывать» или «настраивать» на объект, но и в этом случае возможно лишь вероятностное управление с коррекцией эффекта обратными связями. Это примерно то же, что делает человеческий разум в процессе любого функционального акта. Разница лишь в степени сложности управляемых объектов и в вероятности эффекта управления.

Действующие модели — аппарат внешней памяти будущего. Они должны заменить библиотеки книг. Однако это не имеет прямого отношения к проблеме интеллекта.

Искусственный интеллект выше человеческого разума

Каким можно представить себе такой искусственный интеллект?

Уже говорилось, что мыслимы различные интеллекты — неодинаковой «мощности» и направленности. Направленность я представляю как градации от универсального к специализированному интеллекту. Разница выражается прежде всего в критериях. Универсальный интеллект создается по типу человеческого:

он отражает потребности «тела», разума, среды, в том числе среды социальной — общества, при значительной самоорганизации в смысле возможности самовоспитания. Специализированный направлен на оптимальное управление определенной сложной системой, и его главные критерии диктуются именно ею. «Личные» качества нужно ограничить, так же как и «воспитуемость». Подобный интеллект наиболее приближается к традиционному понятию робота. Например, он сможет присматривать за маленьким ребенком. Искусственный интеллект такого направления при высоком уровне сознания должен быть личностью, поскольку ему придется общаться с людьми. Однако требование преимущества главной потребности — стремления к «благу» управляемого сложного объекта — должно оставаться непреложным. Здесь вступает в силу главный «закон робототехники»: не вредить людям. Сомнительно лишь, возможно ли его соблюсти, поскольку высокий уровень ИИ предусматривает творчество и способность к перевоплощению. Он должен уметь создавать новые методы управления в связи с изменением объекта и обстановки. Не придумает ли он и новые убеждения? Можно ли найти ограничители, способные удержать его в положении специалиста-служаки, живущего одними только интересами дела?

Универсальный интеллект высокого уровня подобен очень умному человеку. Думаю, что у него должен быть тот же принцип действия: сознание, подсознание через СУТ, обобщение моделей разного уровня, ФА. Реализация всего этого зависит от технологии. От того, будет ли это чисто алгоритмический интеллект или в него заложат элементы сетевого интеллекта на физических элементах, зависит многое.

Важнейший вопрос — выбор потребностей (критериев, чувств); видимо, нужен такой же набор их, как и у человека. Вся трудность — в выборе характеристик, значимости, в определении возможности ограничений и воспитуемости. При высоком уровне сознания убеждения приобретают главенствующее значение, однако «врожденные» потребности в большой степени их направляют. Сразу возникает сомнение: а не будет ли ИИ способен регулировать характеристики этих «врожденных» потребностей? Человек не может этого делать, и воспитуемость его в зрелом возрасте весьма ограничена. Но у ИИ будут другие возможности! Этот вопрос об ограничителях, иными словами, та же

модифицированная «робототехника» остается самым важным.

Не буду останавливаться на «характере» ИИ, он связан с теми же проблемами характеристик.

Речь у ИИ, разумеется, будет представлена, причем в нескольких вариантах: одна — для общения с людьми и книгами, другая — для связей с подобными себе. Эта последняя система знаков может быть более совершенной и «технологической».

Творчество — вот главная цель создания искусственного интеллекта уровня выше человеческого разума. Как было сказано, простые программы создания новых моделей представляют собой перебор моделей низшего уровня по обобщенным моделям, которыми выражена задача. Успех такого «конструкторского» творчества определяется эрудицией — набором имеющихся в памяти моделей-деталей для обобщенных моделей. Для творчества высокого порядка характерно наличие только самых общих моделей, описывающих объект — будущую машину, которую нужно изобрести, или сложную систему, работу которой, например, нужно объяснить. Для этого приходится привлекать данные из других областей науки, с тем чтобы заполнять предположениями большие пробелы между «островками» отдельных фактов. Можно полагать, что ИИ высокого уровня в этом отношении превзойдет человеческий разум, поскольку у него будет больше знаний и совершенная система подпрограмм подсознательного поиска данных в разных областях науки. Я совсем не преуменьшаю трудности такого поиска даже при очень совершенной организации памяти, особенно если нужно собрать целую цепь моделей для создания гипотезы, объясняющей работу сложной системы. Тем не менее они преодолимы. При поиске нового человеку трудно выйти из узкого круга привычных истин просто потому, что это не позволяют хорошо натренированные модели и связи, которые и представляют собой данную область науки. Даже подсознательный поиск не помогает — за редким исключением гениев. Искусственный интеллект может себе позволить «раскованное мышление», и его творчество будет более эффективным. Впрочем, в этом таятся свои опасности, но не станем снова вдаваться в фантастику. Еще очень далеко до такого интеллекта!



Проект алгоритмической модели интеллекта

Как же практически подойти к построению алгоритмической модели? Согласно правилам эвристического моделирования прежде всего нужно определить ее назначение, выбрать цель. Смоделировать человеческий разум — это кажется столь же просто, сколь и невозможно. Нужен компромисс. Минимально доказательные человеческие качества интеллекта — это речь с перевоплощением в собеседника с образным и словесным мышлением, это третий уровень сознания — слежение за собственными мыслями. Наши предыдущие модели этих качеств не имели, попытки воспроизвести речевое (вербальное) поведение были, но они совершались в отрыве от других программ интеллекта.

Для того чтобы «вместить» эти многообещающие задачи в модель и достигнуть демонстративности, необходимо наметить ограничения. Прежде всего они определяются сюжетом. Я предполагаю воспользоваться привычным для нас путешествием по некоей искусственной среде, хотя можно было бы взять задачу моделирования другого вида деятельности, например работу врача по диагностике и лечению больного или строителя, создающего конструкции из элементов. Предполагается по сюжету, что «субъект» должен иметь главную цель — дойти по компасу и найти предмет — какие-нибудь «три дерева», чтобы обнаружить там «награду», «плату», действующую на чувство

собственности. По пути он преодолевает умеренные трудности, выбирает маршрут, мобилизуя иерархию ФА, думает, используя внутреннюю речь, и периодически общается по радио с партнерами, к которым питает положительные чувства и в личности которых «перевоплощается». Важно, чтобы разнообразие «входов» среды было минимальным, чтобы число уровней усложнения объектов, за которыми следуют уровни обобщения моделей, было ограничено двумя-тремя. Например: модели-образы деревьев (ветки, листья, высокие, средние, низкие, густые, редкие) должны характеризовать лес.

Минимизация разнообразия должна касаться всех элементов модели. К примеру, такой минимум действий: идти — шагать — быстро, медленно, сидеть, есть, разговаривать, думать, вспоминать. Речь придется ограничить ответами на вопросы и, может быть, короткими рассказами.

Модели действий будут разбиты по степени обобщенности на несколько уровней. Наверху — обобщенные действия, которые соответствуют желаниям: «напрягаться» — «расслабиться», «двигаться» — «отдыхать», «самовыражаться» (в смысле «рассказывать») — «замкнуться», «устраняться», «избегать» — «сопротивляться» и т. п.

Первичных желаний будет немного. Так же придется ограничить и расшифровку желаний в конкретные действия — от обобщенных до детальных.

Восприятие будет ограничено условным рецептором зрения с минимальной настройкой по направлению, глубине и напряжению. То, что попадает в поле зрения, станет автоматически кодироваться цифровым шифром: «главная фигура» и «фон». Рецептор слуха будет действовать только для речи.

Чувства-потребности являются важнейшим элементом любого интеллекта. Минимальный их набор примерно такой: голод, боль, страх, любопытство, потребность действовать, утомление и скука, свобода, отношение (симпатия) к партнеру по разговору, самовыражение. Убеждения будут представлены долгом и волей: словесными формулами, диктующими, как нужно поступать в тех или иных случаях. Разумеется, будут универсальные чувства *Пр* — *НПр*, которые и являются главным критерием для выбора действия из нескольких возможных вариантов. Их соотношение определяет уровень душевного комфорта. Эмоции мне

представляются как крайнее выражение чувств: радость — при самом высоком уровне *Пр*, горе — при самом высоком уровне *НПр*. Гнев, агрессивность — это ответ на угрозу, на ограничение свободы. Страх, ужас — пассивная эмоция как ответ на подавляющую угрозу. Каждая эмоция имеет свое желание — обобщенное действие. Гнев, например, вызывает желание оказать сопротивление источнику угрозы. Вопрос об эмоциях требует еще дополнительной проработки. Характер вырисовывается в соотношениях *Пр* — *НПр* и характеристиках чувств и СУТ.

Кроме перечисленных чувств, будут еще дополнительные критерии, о которых уже шла речь. Первый критерий — обобщенность модели, находящейся в сознании, второй, используемый при сравнении моделей, — вероятность, третий — коэффициент времени для оценки значимости будущих событий и четвертый — реальность.

Человек, личность имеет прошлое, оно постоянно присутствует в настоящем. Биографию придется придать и нашему «субъекту». Пока трудно определить ее объем, но он должен быть минимально необходимым.

Ограничения по времени выразятся в продолжительности временного такта — что-нибудь около 5—10 секунд реального времени. Длительность исследования модели, то есть путь, который будет пройден по местности-карте, будет целиком зависеть от возможностей программ, компьютеров и настойчивости экспериментатора. Отрезок времени должен дать доказательную информацию о «человекоподобии» модели.

Схема интеллекта в самом общем виде показана на рис. 4, 21. Поскольку это алгоритм, а не сеть, то отражать на схеме связи, видимо, нет смысла, так как получится многомерная структура. При создании модели она понадобится, но будет столь сложна, что окажется непригодной для восприятия. Тем не менее нужна какая-то систематизация. Поэтому модели, выполняющие одинаковые функции, придется объединить в сферы, которые явятся в то же время и «координатами» сознания. Примерный перечень сфер таков:

1. «Входы» — модели объектов среды.
2. «Выходы» — модели действий. Сюда войдут и модели настройки рецептора.
3. Чувства — модели всех перечисленных критериев, кроме последней группы дополнительных. Каждое чувство должно иметь свой отдел в сфере.

4. Модель самого себя и отношений к среде. Складывается из ощущений, получаемых с различных следящих систем — с рецепторов, воспринимающих внешнюю среду; с рецепторов «тела», дающих некоторые чувства, со следящей системы, регистрирующей переключения СУТ, то есть наблюдающей за своими действиями и моделями.

5. Комплекс моделей собеседника с его чувствами, действиями и предполагаемыми мыслями, то есть то, что реализует программу перевоплощения. Сочетание этой сферы с предыдущей составят «чувства соперживания» — компоненты собственных чувств, вызванных чувствами собеседника.

6. Независимая координата времени. Она должна иметь свои модели и свою сферу. Ее разделы — настоящее, прошедшее, будущее.

7. Отдельная сфера для критериев вероятности, реальности, обобщенности. Это важные критерии — координаты для любой «вещественной» модели.

8. Сфера программ переключения этапов функционального акта. Ее модели осуществляют слежение за выполнением алгоритма и этапов ФА — восприятие, анализ, планирование, решение, действия.

Думаю, что не нужна отдельная сфера «речь», просто потому что модели слов будут содержаться в каждой сфере в качестве отдельного кода, наряду с моделями-образами.

Сферы в то же время предполагают определенную организацию памяти, как кратковременной, так и внешней. Но сначала поговорим о «кирпичиках» памяти — элементарных моделях. Мне представляется, что для модели данного интеллекта нужно задать «словарь слов», то есть перечень элементарных моделей — понятий, образов и слов речи — для каждой сферы, с делением их по степени обобщенности. Иначе говоря, создать иерархически построенную систему моделей. Приблизительная прикидка такой системы приведена для сферы «выходы — действия». Индекс степени обобщенности должен присутствовать в модели (рис. 42). Модель низкого уровня (детальная) следует снабдить адресом, указывающим на принадлежность ее к более высокому классу понятий.

Модель «шагать с усилием» должна включать цифры, указывающие, что это значит «идти», что это «движение», «действие». Обобщенная по классу I модель «действие» предполагает «вакантные» места для

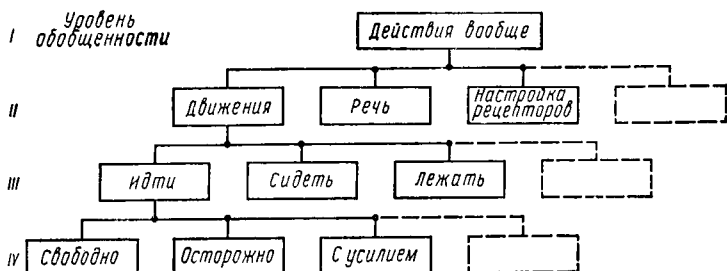


Рис. 42. Система моделей «Выходы».

понятий следующего уровня обобщенности — как бы адреса соответствующих моделей.

Таким же образом строятся системы моделей для понятия «предметы» (имеется в виду внешняя среда). В подробности сейчас я не вдаюсь просто потому, что работа над моделью только начата и все еще очень неопределенно.

Элементарные модели разной степени обобщенности составляют «словари слов» в каждой сфере. Однако основой мышления являются не «слова», а «фразы», причем «фразы» короткие, не более трех слов. «Словари фраз» могут составляться из «слов» данной сферы, например между моделями разных уровней обобщенности («движение» — «лежать», «движение» — «сидеть» или, наоборот, «идти» — «движение»). То же относится к соединению моделей-образов с моделями слов речи (образ камня и слово «камень»).

Однако самые важные «фразы» объединяют «слова», принадлежащие к разным сферам, например «предмет — чувство» или «чувство — действие». Видимо, между каждыми двумя сферами должно существовать пограничное поле, наполненное такими «фразами» поскольку связи между ними имеют одностороннюю направленность. Мало того, при связях должен еще быть «коэффициент проходимости», указывающий на степень сродства двух понятий. Можно представить себе довольно много «словарей фраз» и еще большее число возможных сочетаний «слов» во «фразах». В этом заключается вся трудность: придется ограничивать число «слов», иначе модель быстро сделается необозримой.

Внешняя память вся представлена «словарями фраз», составленными так, что на первом месте стоят

наиболее употребительные, на последнем — редко встречающиеся «фразы».

Мне трудно сейчас определить состав кратковременной или оперативной памяти. В ней должны находиться модели, являющиеся «координатами» сознания. Это прежде всего модели из каждой сферы, поскольку здесь представлено слежение, осуществляемое всеми типами рецепторов. Зрение определяет «субъекта» в пространстве, «служба времени» отмечает, какое сейчас время, рецепторы «тела» говорят об ощущениях и чувствах, так же как рецепторы самого интеллекта — например следящие за переключением сознания. В оперативной памяти каждой сферы будет сколько-то активных моделей, которые получены рецепторами от объектов своего слежения. Чем более активная сфера, то есть чем более напряжен соответствующий рецептор, тем больше будет моделей, тем они будут активнее. «Старые» модели станут частью забываться, частью переводиться во внешнюю память в том случае, если они долго хранились в оперативной памяти и повторно привлекались в сознание. Другим источником моделей для оперативной памяти является память внешняя. По алгоритму ФА будут извлекаться модели — «фраза» по их первому «слову» для вспоминания, сравнения, прогнозирования. Подробности этого механизма еще не прояснились.

Расчет активности моделей в оперативной памяти является важнейшей операцией алгоритмического интеллекта, поскольку активность моделей определяет движение сознания. Параметры активности всех моделей оперативной памяти пересчитываются в каждый такт времени. Для расчета служат характеристики — статические и динамические, по типу показанных на рис. 6 и 7. Самым трудным явится расчет циркуляции энергии по связям между моделями оперативной памяти, от которого зависит их активность. Все подсознание зиждется на таком расчете. По всей вероятности, для этого придется привлекать коэффициенты проходимости связей, зафиксированные в «словарях фраз» внешней памяти. Например, если в оперативной памяти есть слово «волк» и чувство «страх», введенные из разных источников, то они будут влиять друг на друга по принципу: более активное слово — на менее активное, а степень влияния определится коэффициентом связи.

Вызов нового «слова» из внешней памяти нужно связать с сознанием, и этот вопрос остается неясным, пока не определится структура СУТ. По всей вероятности, нужна двухуровневая СУТ, какую мы уже использовали в МОДе, — «микро-СУТ» отдельно для каждой сферы и «макро-СУТ, выбирающая максимально активную сферу. Возможно, «микро-СУТ» будет дано право вызывать из памяти модели, с тем чтобы осуществлять простые ФА в подсознании.

Общий алгоритм действия АИ будет построен на функциональных актах. Все движение по местности и общение с собеседником должны представлять иерархию и сеть ФА, из которых несколько останутся незаконченными и составят содержание мыслей. Во время отдыха или движения по легкой местности «субъект» может «отвлекаться» — на воспоминания, нереальное планирование. Предметом воспоминаний будут активные модели в подсознании, способные привлечь СУТ в интервалах ФА. Утомление и скука являются важным регулятором продолжительности однотипных действий.

Внутренняя речь будет сопровождать действия с моделями-образами. Общение по радио, а возможно, какая-нибудь иная его форма, еще требует проработки. Видимо, диалог будет представлять самостоятельные ФА; вкрапленные среди других.

Вот главные трудности на пути создания модели интеллекта:

1. Орган зрения. На сетчатке глаза миллионы рецепторных клеток и еще несколько слоев вспомогательных нейронов. Это не просто миниатюрные датчики, это их система, воспринимающая пространственную картину и выделяющая ее некоторые свойства, то есть несущая функции первичного распознавания. Воспроизвести глаз техническими средствами пока не удастся. Не случайно проблема распознавания образов в кибернетике — одна из самых трудных. В предполагаемой модели АИ мы попытаемся обойти ее, сразу вводя образы в кратковременную память в цифровом выражении. Разумеется, это возможно только при очень упрощенной среде и ограниченном наборе объектов.

2. Представление мира-среды и самого себя моделями разной степени обобщенности — от детальных картин до самых общих понятий. В памяти должно быть несколько моделей одной и той же картины или ее изменений во времени. Нужен удобный механизм

перехода от детальной картины к обобщенной, и наоборот. Мы предполагаем обойти эту трудность созданием нескольких дискретных уровней обобщения моделей и алгоритма перехода между ними.

3. Подсознание, изменение активности всех моделей, находящихся в памяти, взаимодействие между моделями, приторможенными СУТ, вплоть до подсознательных двигательных актов. Придется ограничить подсознание областью кратковременной памяти, выделив отдельно внешнюю память, в которой модели не будут изменять свою активность. Параллельный пересчет активности можно осуществить алгоритмически.

4. Самоорганизация. Она выражается в образовании новых связей и повышении собственной активности моделей в результате тренировки. Это качество можно воспроизвести при относительно небольшом объеме информации.

Едва ли стоит продолжать предположения и планы, касающиеся будущей модели. Все может измениться в процессе работы. Будет модель — будут и описания, а теперь нужно становиться.

з а к л ю ч е н и е

Попытаемся подытожить основные идеи гипотезы об алгоритме разума.

Интеллект — это аппарат управления сложными системами через действия с их моделями для достижения максимума критериев оптимальности. Понятие интеллекта применимо к органам управления всех сложных систем «типа живых» — ДНК в клетке, нервная система в организме, администрация в обществе. Интеллект может быть воплощен различными материальными средствами от биологических до технических.

Типовая схема интеллекта включает обязательные элементы — рецепторы, эффекторы и «мозг», содержащий модели среды, «тела», критериев и действий.

Объектами управления интеллекта могут быть собственное «тело», внешний мир, другие разумные существа и, наконец, он сам — разум. Каждый из них представляет собственные критерии оптимальности в виде моделей с высокой активностью, содержащих компоненты универсального критерия «Приятно — Неприятно». Деятельность разума направлена на получение максимального уровня душевного комфорта (УДК), то есть максимума суммы $\Sigma (Пр - НПр)$.

Основная проблема интеллекта — преодоление избыточности разнообразия внешнего мира, его восприятия множеством рецепторов, его картины во временной памяти, вариантов возможных действий. Нужно выбрать одно действие, удовлетворяющее многим критериям в настоящем и будущем. Это возможно только при использовании моделей разной обобщенности. Модели должны отражать пространственную структуру внешнего мира и своего тела, различные виды сил и энергий как внешних, так и собственных, критерии раз-

ной значимости и, наконец, изменение всего этого во времени.

Суть алгоритма интеллекта состоит в том, чтобы от множества активных моделей «входов» и критериев, с учетом их прошлого и прогноза на будущее, выбрать и активировать одну модель действий. Для этого модели должны обладать следующими параметрами:

а) сложной структурой, отражающей объекты с разной степенью обобщенности;

б) активностью (потенциалом энергии), изменяющейся по статическим и динамическим характеристикам в зависимости от энергии, получаемой по связям от рецепторов и других моделей. На этом основана кратковременная память;

в) свойством тренируемости, то есть изменением характеристик в зависимости от интенсивности и длительности предшествовавшей активности;

г) способностью образовывать новые связи между наиболее активными моделями и тренировать их при частом использовании.

Тренировка моделей и связей обеспечивает обучение, воспитание и самоорганизацию разума.

Важнейшими действиями с моделями являются: активирование новой модели от других, сравнение моделей для определения общего и различий, превращение частных моделей в обобщенную и, наоборот, развертывание обобщенной модели в набор частных.

Эффективное управление, то есть выбор действий и доведение их до цели при наличии «помех» со стороны других раздражителей среды и критериев, возможно только при реализации специального алгоритма усиления моделей наиболее значимого функционального акта и торможения всех других ФА. Такой алгоритм возможен в виде системы усиления-торможения — СУТ. При этом модельно воспроизводятся основные психологические понятия — сознание, подсознание, мысль.

Эволюция «мощности» интеллекта выражается в возрастании способности отвечать на максимальное и изменяющееся разнообразие среды действиями, спланированными и реализованными во времени так, чтобы получать максимум эффекта по многим критериям в длительные отрезки времени. Это возможно путем взаимодействия ФА разной обобщенности и направленности при возрастающем количестве и сложности моделей в памяти.

Можно условно выделить несколько ступеней эволюции интеллектов:

Первая — жесткий автомат. Структурно это прямая связь: рецептор → усилитель → эффектор. Функционально: признак в среде → действие. Критерии заложены в структуре связей и специфике рецептора. ФА короткий, без обратных связей, с блокировкой до окончания действия.

Вторая — сложный многопрограммный автомат. В структуре — несколько рецепторов и эффекторов. Разум представлен сложной системой коммуникаций, включающей обратные связи, усилители, блокировку, торможение. Функция выражается в способности выделять несколько видов признаков и в зависимости от набора включать один из нескольких ФА, состоящих из последовательности действий, выполняемых под контролем обратных связей, при блокировке всех других ФА. Алгоритм реализован в клетках и технике.

Третья — разум животных, не имеющих коры головного мозга. Он похож на сложный автомат. Отличие: много рецепторов, вероятностное распознавание — выбор моделей из постоянной памяти, включение моделей действий в зависимости от критериев состояния внутренней среды. От них же могут включаться «ФА поиска». Ограниченное обучение реализуется за счет избирательной тренировки нейронов, избыточных их связей и мышц. Выделение самого значимого ФА обеспечивается торможением, блокировкой всех других, но при этом возможно слежение за средой для экстренного переключения на другой ФА.

Четвертая — разум млекопитающих, представленный корой головного мозга, этажом над подкоркой. Кора состоит из большого числа нейронов с удлиненными характеристиками, обеспечивающими сложные модели в памяти с разной степенью обобщенности. Функционируют СУТ, элементарное сознание, система критериев, представлены все этапы ФА, включая короткое прогнозирование и планирование. Обучение достигается между нейронами. Возможно случайное творчество через запоминание удачных проб, заложенных в «ФА поиска». Специальной дрессировкой (тренировкой) человек может привить животному простейшие «убеждения».

Пятая — разум человека. Его преимущества, видимо, объясняются большим количеством нейронов «новой коры» с удлиненными характеристиками и повы-

шенной тренируемостью. Это привело к созданию сложных моделей, большому объему памяти и высокой собственной активности новых моделей, достигаемой упражнением. Отсюда все следствия: обобщенные модели, охватывающие большие участки пространства и интервалы времени, речь, через нее — воспитание и обучение, и в частности самоорганизация и формирование убеждений — активных речевых моделей поведения. Высокий уровень сознания выражается в слежении за мыслями, перевоплощении, овладении «нереальным планированием» — мечтами, искусством. Творчество и труд, как создание новых моделей и их физическое воплощение, являются алгоритмом развертывания обобщенных моделей задачи в детальные и считывания их действиями.

Шестая — искусственный интеллект выше уровня человека. Если не принимать во внимание трудности технологии, то можно предположить моделирование человеческого разума с увеличением его собственной памяти и созданием нового типа памяти внешней в виде количественных действующих моделей, призванных дополнить библиотеки из книг. Это расширит возможности творчества и самопознания, что и явится самым главным продуктом интеллекта.

хв хв хв хв

*

Чувствую, что книга вышла несовершенной. В ней нет ссылок и цитат, мало имен авторитетов, которые должны подкрепить высказывания автора. Мне не хотелось перегружать изложение нарочитой «научностью», чтобы не отвлекать внимание от основной идеи. Тем более что книга рассчитана на читателей очень разных специальностей — от инженеров до врачей. В конце концов я хотел лишь представить новый подход к пониманию механизмов психики — не от физиологического анализа, а от кибернетического синтеза. К сожалению, такой подход имеет важный недостаток: истину нужно доказывать работающей моделью. Пока ее нет, все построения висят в воздухе.

Может быть, не следовало торопиться писать? Подождать, пока будет модель? Но, ох как медленно они делаются, модели! Поэтому прошу рассматривать эту книгу как гипотезу. Они ведь тоже иногда приносят пользу.

Список литературы

1. Амосов Н. М. Моделирование мышления и психики.— Киев, 1965.
2. Амосов Н. М. Моделирование процессов мышления.— Кибернетика, 1968, № 2.
3. Амосов Н. М. Искусственный разум.— Киев, 1969.
4. Амосов Н. М., Касаткин А. М., Касаткина Л. М., Талаев С. А. Автоматы и разумное поведение: Опыт моделирования.— Киев, 1973.
5. Виноград Т. Программа, понимающая естественный язык.— М., 1976.
6. Касаткина Л. М., Касаткин А. Эвристическая модель поведения.— В кн.: Некоторые проблемы биоккибернетики... Киев, 1966, вып. 2, с. 21—36.
7. Мак-Каллох У. Надежность биологических систем.— В кн.: Самоорганизующиеся системы. М., 1964, с. 358—378.
8. Минский М., Пейперт С. Перцептроны.— М., 1971.
9. Сутро Л., Киллмер У. Совокупность решающих устройств для управления роботом.— В кн.: Интегр. роботы. М., 1973, с. 112—163.
10. Нильсон Н. Искусственный интеллект.— М., 1973.
11. Сэмюэль А. Некоторые исследования возможности обучения машин на примере игры в шашки.— В кн.: Вычислительные машины и мышление. М., 1967, с. 71—112.
12. Ньюэл А., Саймон Г. GPS — программа, моделирующая процесс человеческого мышления.— Там же, с. 283—301.
13. Шенк Р., Абельсон Р. Сценарий, планы и знание.— В кн.: Тр. IV Междунар. объедин. конф. по ИИ. М., 1975, с. 208—220.
14. Файкс Р., Нильсон Н. Система STRIPS ...— В кн.: Интегр. роботы. М., 1973, с. 382—403.
15. Hewitt C. PLANNER language for proving Theorems in Robots.— In: Int. joint. conf. Art., Intel. Wash., 1969, p. 295—301.
16. Hewitt C. Procedural Imbedding of Knowledge in PLANNER.— In: Second Int. joint. conf. Art. Intel. L. 1971, p. 167—182.
17. Robinson J. A. The generalized resolution principle. N.-Y., 1968, 3, p. 77—94.

Содержание

Предисловие	5
Введение	10
Искусственный интеллект сегодня	14
Основные этапы и направления исследований	14
Некоторые результаты моделирования сетевого интеллекта	18
Системы и модели	31
Сложные системы	33
Модели	35
Восприятие и память	38
Восприятие и управление	38
Обобщенность восприятия	40
Виды и механизмы памяти	43
Гипотеза о механизмах памяти	46
Вспоминание, обобщение, забывание	50
О реализации гипотезы	53
Действия с моделями	55
Активация моделей	55
Сравнение моделей и распознавание образов	57
Дописывание «фразы» — вспоминание	60
Обобщение моделей	60
Взаимодействие моделей в интеллекте	63
Критерии, потребности, чувства, стимулы	66
Классификация потребностей-чувств	73
Такты деятельности и напряжение	78
Время	80
Реальность	82

Функциональный акт	85
Алгоритм упрощенного функционально-го акта	87
«Круги» восприятия	95
Прогнозирование и определение динамики	97
Иерархия функциональных актов	99
Сеть функциональных актов	103
«Мысли»	106
Сознание и подсознание	109
Система усиления-торможения — СУТ	109
СУТ в алгоритмическом интеллекте	111
СУТ и сознание	114
Функции подсознания	116
«Координаты» сознания	120
Интеллект уровня человека	122
Речь	123
Высшие уровни сознания	149
Сопереживание и воображение	153
Труд и творчество	157
Сновидения и психологические болезни	164
Искусственный интеллект в человеческом обществе	170
Основные «индивидуальные» качества ИИ	170
Общество искусственных интеллектов	188
На пути к интеллекту выше человеческого	195
Разум человека и сложные системы	195
Метод эвристического моделирования	199
Эвристические модели некоторых сложных систем	202
Искусственный интеллект выше человеческого разума	206
Проект алгоритмической модели интеллекта	209
Заключение	216
Список литературы	221

Николай
Михайлович
АМОСОВ

АЛГОРИТМЫ РАЗУМА

Печатается по постановлению
Редакционной коллегии
научно-популярной литературы
АН УССР

Заведующий
редакцией
А. М. Азаров

Редактор
С. М. Хазанет

Художественный
редактор
В. И. Прищепа

Оформление
художника
Р. К. Пахолюка

Технический
редактор
И. Н. Лукашенко

Корректоры
О. Е. Исарова,
В. Н. Божок,
М. В. Гайдамак

Информ. бланк № 2240

Сдано в набор 29.01.79.

Подп. в печ. 17.07.79.

БФ 01640.

Формат 84×108/32.

Бумага типогр. № 1. Школьн. гарн.

Выс. печ. Усл. печ. л. 11,76. Уч.-изд. л. 11,5.

Тираж 37 000 экз. Заказ № 9—203.

Цена 40 коп.

Издательство «Наукова думка»,
252601, Киев, ГСП, Репина, 3.

Напечатано с матриц Головного
предприятия республиканского производственного
объединения «Полиграфкнига»
Госкомиздата УССР,
252057, г. Киев-57, Довженко, 3
в Киевской книжной типографии
научной книги, 252004, Киев-4, Репина, 4.
Зак. 9—777.