

## Лекция 5. Архитектура и основные составные части систем ИИ.

Различные подходы к построению систем ИИ (логический, структурный, эволюционный, имитационный) и методы представления знаний. Краткое ознакомление с данными подходами. Вспомогательные системы (распознавание образов зрительных и звуковых, идентификация, моделирование, жесткое программирование) и их место в системах ИИ.

Лекционные занятия - 2 ч.

### Различные подходы к построению систем ИИ

Существуют различные подходы к построению систем ИИ. Это разделение не является историческим, когда одно мнение постепенно сменяет другое, и различные подходы существуют и сейчас. Кроме того, поскольку по-настоящему полных систем ИИ в настоящее время нет, то нельзя сказать, что какой-то подход является правильным, а какой-то ошибочным.

Для начала кратко рассмотрим логический подход. Почему он возник? Ведь человек занимается отнюдь не только логическими измышлениями. Это высказывание конечно верно, но именно способность к логическому мышлению очень сильно отличает человека от животных.

Основой для данного логического подхода служит Булева алгебра. Каждый программист знаком с нею и с логическими операторами с тех пор, когда он осваивал оператор IF. Свое дальнейшее развитие Булева алгебра получила в виде исчисления предикатов — в котором она расширена за счет введения предметных символов, отношений между ними, кванторов существования и всеобщности. Практически каждая система ИИ, построенная на логическом принципе, представляет собой машину доказательства теорем. При этом исходные данные хранятся в базе данных в виде аксиом, правила логического вывода как отношения между ними. Кроме того, каждая такая машина имеет блок генерации цели, и система вывода пытается доказать данную цель как теорему. Если цель доказана, то трассировка примененных правил позволяет получить цепочку действий, необходимых для реализации поставленной цели. Мощность такой системы определяется возможностями генератора целей и машиной доказательства теорем.

Конечно можно сказать, что выразительности алгебры высказываний не хватит для полноценной реализации ИИ, но стоит вспомнить, что основой всех существующих ЭВМ является бит — ячейка памяти, которая может принимать значения только 0 и 1. Таким образом было бы логично предположить, что все, что возможно реализовать на ЭВМ, можно было бы реализовать и в виде логики предикатов. Хотя здесь ничего не говорится о том, за какое время.

Добиться большей выразительности логическому подходу позволяет такое сравнительно новое направление, как нечеткая логика. Основным ее отличием является то, что правдивость высказывания может принимать в ней кроме да/нет (1/0) еще и промежуточные значения — не знаю (0.5), пациент скорее жив, чем мертв (0.75), пациент скорее мертв, чем жив (0.25). Данный подход больше похож на мышление человека, поскольку он на вопросы редко отвечает только да или нет. Хотя правда на экзамене будут приниматься только ответы из разряда классической булевой алгебры.

Для большинства логических методов характерна большая трудоемкость, поскольку во время поиска доказательства возможен полный перебор вариантов. Поэтому данный подход требует эффективной реализации вычислительного процесса, и хорошая работа обычно гарантируется при сравнительно небольшом размере базы данных.

Под **структурным подходом** мы подразумеваем здесь попытки построения ИИ путем моделирования структуры человеческого мозга. Одной из первых таких попыток был перцептрон Френка Розенблатта. Основной моделируемой структурной единицей в перцептронах (как и в большинстве других вариантов моделирования мозга) является нейрон.

Позднее возникли и другие модели, которые в простонародье обычно известны под термином "нейронные сети" (НС). Эти модели различаются по строению отдельных нейронов, по топологии связей между ними и по алгоритмам обучения. Среди наиболее известных сейчас вариантов НС можно назвать НС с обратным распространением ошибки, сети Хопфилда, стохастические нейронные сети.

НС наиболее успешно применяются в задачах распознавания образов, в том числе сильно зашумленных, однако имеются и примеры успешного применения их для построения собственно систем ИИ, это уже ранее упоминавшийся ТАИР.

Для моделей, построенных по мотивам человеческого мозга характерна не слишком большая выразительность, легкое распараллеливание алгоритмов, и связанная с этим высокая производительность параллельно реализованных НС. Также для таких сетей характерно одно свойство, которое очень сближает их с человеческим мозгом — нейронные сети работают даже при условии неполной информации об окружающей среде, то есть как и человек, они на вопросы могут отвечать не только "да" и "нет" но и "не знаю точно, но скорее да".

Довольно большое распространение получил и **эволюционный подход**. При построении систем ИИ по данному подходу основное внимание уделяется построению начальной модели, и правилам, по которым она может изменяться (эволюционировать). Причем модель может быть составлена по самым различным методам, это может быть и НС и набор логических правил и любая другая модель. После этого мы включаем компьютер и он, на основании проверки моделей отбирает самые лучшие из них, на основании которых по самым различным правилам генерируются новые модели, из которых опять выбираются самые лучшие и т. д.

В принципе можно сказать, что эволюционных моделей как таковых не существует, существует только эволюционные алгоритмы обучения, но модели, полученные при эволюционном подходе имеют некоторые характерные особенности, что позволяет выделить их в отдельный класс.

Таковыми особенностями являются перенесение основной работы разработчика с построения модели на алгоритм ее модификации и то, что полученные модели практически не сопутствуют извлечению новых знаний о среде, окружающей систему ИИ, то есть она становится как бы вещью в себе.

Еще один широко используемый подход к построению систем ИИ — **имитационный**. Данный подход является классическим для кибернетики с одним из ее базовых понятий — "черным ящиком" (ЧЯ). ЧЯ — устройство,

программный модуль или набор данных, информация о внутренней структуре и содержании которых отсутствуют полностью, но известны спецификации входных и выходных данных. Объект, поведение которого имитируется, как раз и представляет собой такой "черный ящик". Нам не важно, что у него и у модели внутри и как он функционирует, главное, чтобы наша модель в аналогичных ситуациях вела себя точно так же.

Таким образом здесь моделируется другое свойство человека — способность копировать то, что делают другие, не вдаваясь в подробности, зачем это нужно. Зачастую эта способность экономит ему массу времени, особенно в начале его жизни.

Основным недостатком имитационного подхода также является низкая информационная способность большинства моделей, построенных с его помощью.

С ЧЯ связана одна очень интересная идея. Кто бы хотел жить вечно? Я думаю, что почти все ответят на этот вопрос "я".

Представим себе, что за нами наблюдает какое-то устройство, которое следит за тем, что в каких ситуациях мы делаем, говорим. Наблюдение идет за величинами, которые поступают к нам на вход (зрение, слух, вкус, тактильные, вестибулярные и т. д.) и за величинами, которые выходят от нас (речь, движение и др.). Таким образом человек выступает здесь как типичный ЧЯ.

Далее это устройство пытается отстроить какую-то модель таким образом, чтобы при определенных сигналах на входе человека, она выдавала на выходе те же данные, что и человек. Если данная затея будет когда-нибудь реализована, то для всех посторонних наблюдателей такая модель будет той же личностью, что и реальный человек. А после его смерти она, будет высказывать те мысли, которые предположительно высказывал бы и смоделированный человек.

Мы можем пойти дальше и скопировать эту модель и получить брата близнеца с точно такими же "мыслями".

Можно сказать, что "это конечно все интересно, но при чем тут я? Ведь эта модель только для других будет являться мной, но внутри ее будет пустота. Копируются только внешние атрибуты, но я после смерти уже не буду думать, мое сознание погаснет (для верующих людей слово "погаснет" необходимо заменить на "покинет этот мир") ". Что ж это так. Но попробуем пойти дальше.

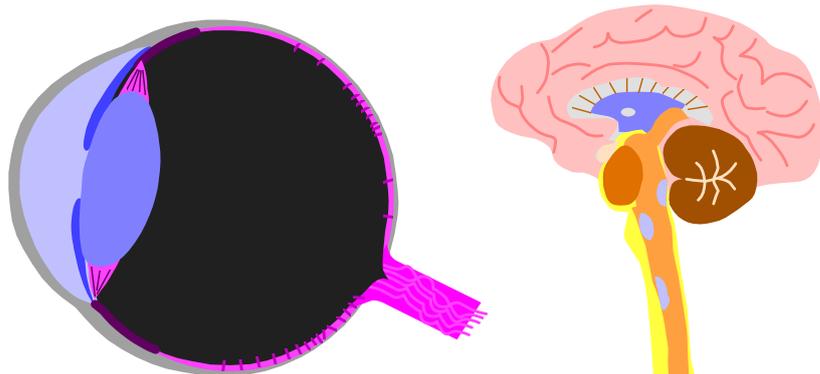
Согласно философским представлениям автора данного курса, сознание представляет собой сравнительно небольшую надстройку над нашим подсознанием, которая следит за активностью некоторых центров головного мозга, таких как центр речи, конечной обработки зрительных образов, после чего "возвращает" эти образы на начальные ступени обработки данной информации. При этом происходит повторная обработка этих образов, мы как бы видим и слышим, что думает наш мозг. При этом появляется возможность мысленного моделирования окружающей действительности при нашем "активном" участии в данном процессе. И именно наш процесс наблюдения за деятельностью этих немногих центров является тем, что мы называем сознанием. Если мы "видим" и "слышим" наши мысли, мы в сознании, если нет, то мы находимся в бессознательном состоянии.

Если бы мы смогли смоделировать работу именно этих немногих "сознательных" нервных центров (работа которых правда основана на деятельности всего остального мозга) в качестве одного ЧЯ, и работу "супервизора" в качестве другого ЧЯ, то можно было бы с уверенностью говорить, что "да, данная модель думает, причем так же, как и я". Здесь я ничего не хочу говорить о том, как получить данные о работе этих нервных центров, поскольку на мой взгляд сегодня нет ничего такого, что позволило бы следить за мозгом человека годами и при этом не мешало бы его работе и жизни.

И заканчивая беглое ознакомление с различными методами и подходами к построению систем ИИ, хотелось бы отметить, что на практике очень четкой границы между ними нет. Очень часто встречаются смешанные системы, где часть работы выполняется по одному типу, а часть по другому.

**Вспомогательные системы нижнего уровня (распознавание образов зрительных и звуковых, идентификация, моделирование, жесткое программирование) и их место в системах ИИ**

Для того, чтобы человек сознательно воспринял информацию (для примера возьмем чертеж), она должна пройти довольно длительный цикл предварительной обработки. Вначале свет попадает в глаз. Пройдя через всю оптическую систему фотоны в конце концов попадают на сетчатку — слой светочувствительных клеток — палочек и колбочек.



Уже здесь — еще очень далеко от головного мозга, происходит первый этап обработки информации, поскольку, например, у млекопитающих, сразу за светочувствительными клетками находится обычно два слоя нервных клеток, которые выполняют сравнительно несложную обработку.

Теперь информация поступает по зрительному нерву в головной мозг человека, в так называемые "зрительные бугры". То, что именно сюда приходит видеоинформация для дальнейшей обработки, показывают многочисленные опыты над людьми во время различных операций, в ходе которых производилась трепанация черепа. При этом пациентам раздражали область зрительных бугров слабым электрическим полем, что вызывало у них различные световые галлюцинации. Причем, что интересно, при изменении места раздражения, пропорционально смещению, смещались и места галлюцинаций, т. е. на зрительные бугры как бы проецируется то, что мы видим.

Некоторые исследователи пошли дальше, и вживляли слепым людям целую матрицу электродов, напряжения на которых соответствовали освещенности соответствующих участков видеокамеры, размещенной на голове пациента. После операции, слепые начинали различать крупные фигуры (квадрат, треугольник,

круг) и даже читать текст (при вживлении матрицы  $10 \times 10$ ). Широкому распространению данного метода лечения слепоты препятствуют как недостаточно высокий наш технический уровень, так и чрезвычайно высокая опасность операций на открытом мозге. Такого рода опыты проводятся только попутно с операцией, вызванной другими причинами.

Далее зрительная информация поступает в отделы мозга, которые уже выделяют из нее отдельные составляющие — горизонтальные, вертикальные, диагональные линии, контуры, области светлого, темного, цветного. До этих пор мы можем без труда смоделировать работу мозга применяя различные графические фильтры. Постепенно образы становятся все более сложными и размытыми, но графический образ картины пройдет еще долгий путь, прежде чем достигнет уровня сознания. Причем на уровне сознания у нас будет не только зрительный образ, к нему примешаются еще и звуки, запахи (если картина представляет собой натюрморт) и вкусовые ощущения. Дальнейшие ассоциации каждый может додумать сам.

Смысл всего сказанного заключается в том, чтобы показать, что в системах ИИ имеются подсистемы, которые мы уже сейчас можем реализовать даже не зная о том, как они реализованы у человека. Причем можем это сделать не хуже, чем у прототипа, а зачастую и лучше. Например, искусственный глаз (а равно и блок первичной обработки видеоинформации, основанные на простейших фильтрах или др. сравнительно несложных устройствах) не устает, может видеть в любом диапазоне волн, легко заменяется на новый, видит при свете звезд.

Устройства обработки звука позволяют улавливать девиацию голоса человека в 1-2 Герца. Данное изменение частоты происходит при повышенном возбуждении вегетативной нервной системы, которое в свою очередь часто обусловлено волнением человека. На данном принципе основаны современные детекторы лжи, которые позволяют обнаружить с высокой вероятностью даже записанные на пленку много лет назад ложные высказывания.

Современные системы управления электродвигателем позволяют с высокой точностью держать заданные координаты даже при ударном изменении нагрузки. А ведь это примерно тоже, что держать на длинной палке баскетбольный мяч, по которому то слева, то справа кидают теннисные мячи.

За одно и тоже время, компьютер произведет гораздо больше арифметических операций и с большей точностью, чем человек.

Антиблокировочная система на автомобилях позволяет держать тормоза на грани заклинивания колеса, что дает наибольшее трение с дорогой, а это без АБС по силам только очень опытным водителям.

В принципе такие примеры, где техника оказывается ничуть не хуже человека, можно продолжать до бесконечности. Общий же смысл сказанного в том, что при конструировании ИИ, мы не связаны одним набором элементарных составляющих, как природа. В каждом конкретном случае желательно применять то, что даст самый большой эффект. В той области, где у человека господствуют рефлексы (чихание, быстрое напряжение быстро растягиваемой мышцы, переваривание пищи, регулировка температуры), мы вообще можем применить жесткие системы управления, с раз и навсегда заданным алгоритмом функционирования. При этом вполне можно ожидать увеличения точности и

уменьшение времени обучения их до нуля. При этом ядро нашей системы ИИ будет решать уже не настолько глобальные задачи.

Данный принцип разбиения задачи на подзадачи уже давно используется природой. К примеру, мы далеко не полностью используем все возможности наших мышц в области разнообразия движений. Мы не можем заставить наши глаза смотреть в разные стороны, не говоря уже о том, чтобы делать это на разном уровне (левый глаз — влево-вверх, правый — вправо-вниз). При ходьбе мы часто используем далеко не оптимальный набор движений и далеко не все сочетания вариантов напряжения мышц мы опробуем. Попробуйте к примеру сделать волну животом. В принципе здесь нет ничего сложного, поскольку каждый пучок мышц пресса иннервируется отдельно, но если Вы этого не делали ранее, то получить необходимый результат будет не просто — в повседневной жизни это действие ненужно, а значит его нет и в "словаре движений", а на обучение необходимо определенное время. А по поводу оптимальности походки существуют расчеты, что если бы человек всегда рассчитывал оптимально траекторию движения в которой существует более 200 степеней свобод, то он бы не ходил, а в основном бы только думал о том, как надо ходить.

На самом деле наша система управления построена по иерархическому принципу, когда задача распределяется между несколькими уровнями. Высший уровень нервной системы (связанный с большими полушариями мозга) ставит лишь общую задачу, скажем, переложить книгу на стол. Этот уровень вообще не контролирует действие отдельных двигательных единиц, направленных на решение поставленной задачи. Здесь уместна аналогия: командующий армией, ставя перед своими войсками некую общую задачу, отнюдь не предписывает каждому солдату и офицеру, что именно он должен делать в каждый момент операции.

Детализация построения движений у человека происходит на уровнях более низких, чем командный уровень коры больших полушарий. Более того, в некоторых случаях (когда мы отдергиваем руку, прикоснувшись к горячему предмету, даже не успев осознать ситуацию) все управление формируется на нижележащих уровнях, связанных с различными отделами спинного мозга.

В общем ситуация схожа с той, когда программист использует библиотеку подпрограмм. При этом ему не важно, какой алгоритм они используют, если программа работает нормально. А написание своей библиотеки тратится драгоценное время. Кроме того, еще не известно, будет ли она работать так же хорошо.

Общий вывод данной лекции состоит в том, что в настоящее время существуют методы, алгоритмы и устройства, которые позволяют нам довольно неплохо смоделировать нижние уровни человеческого интеллекта, причем совсем не обязательно на таком же физическом принципе. И если бы это была не лекция, а тост, то я бы закончил его: "...так выпьем же за протестированные, правильно работающие и бесплатные библиотеки подпрограмм".