ДМК ПРЕСС
издательство

Потопахин В.В.

Романтика искусственного интеллекта

В. В. Потопахин

Романтика искусственого интеллекта

Москва, 2017

УДК 004.8
ББК 32.813
П64

Потопахин В. В.
П64 Романтика искусственного интеллекта. – М.: ДМК Пресс, 2017. – 170 с.: ил.

ISBN 978-5-97060-476-2

Эта книга о том, чем занимаются специалисты по искусственному интеллекту. О том, в решении каких задач умные машины уже заменили человека, и какие интеллектуальные технологии могут появится в обозримом будущем. О том, может ли машина стать равноценным партнером человека или даже превзойти его. Насколько реальна возможность бунта машин, так любимого писателями-фантастами? А может быть искусственный интеллект – это просто область технического моделирования поведения, которое мы считаем разумным? И как понять, что умные машины уже живут рядом с нами?

Издание предназначено для широкого круга читателей, интересующихся вопросами искусственного интеллекта.

УДК 004.8
ББК 32.813

ISBN 978-5-97060-476-2

Содержание

Предисловие

Эта книга о проблеме, которую очень многие ученые считают ключевой для развития человеческих технологий, – проблеме искусственного интеллекта. Таковой она является по той причине, что создание полноценного разума, равнозначного человеческому или даже его превосходящего, может означать, что дальнейшее развитие земных технологий уже может пойти без приставки «человеческие». Саморазвивающийся разум станет, в лучшем для людей случае, нашим полноправным партнером в развитии цивилизации на планете.

Впрочем, пока это область фантастики. Самые лучшие представители систем искусственного интеллекта пока совершенно не обнаруживают ключевых возможностей разума: интуиции, творчества, свободы мышления. Поэтому пока термин «искусственный интеллект» именует собой инженерную дисциплину, цель которой – лишь имитировать, в частных задачах, поведение, которое мы, люди, считаем разумным. Каким образом человек, рождаясь практически чистым от знаний, получает способность обучаться и развивать свой разум, почему мы способны принимать решения при недостатке информации, каковы механизмы творчества – это все вопросы, пока слишком сложные для теории искусственного интеллекта.

Ряд крупных ученых даже полагает, что эта теория так всегда и останется чисто инженерной областью, впрочем, может быть, и наш интеллект – просто сверхсложная инженерная конструкция, но тогда, как говорит Дэвид Вернон, мы, может быть, и не заметим, что полноценный интеллект уже живет среди нас.

В общем, эта книга об интеллекте, проблемах его разработки и критериях его распознания. Таковой критерий сам по себе представляет собой отдельную задачу. Действительно, если мы видим перед собой автомобиль, или воздушный шар, или телескоп, нет необходимости в специальных процедурах высокой степени сложности, чтобы убедиться в том, что мы видим автомобиль или воздушный шар. А вот сложность вопроса, что перед нами: машина, работающая по алгоритму, или мыслящее существо, – неимоверно велика. Это вопрос философский, мировоззренческий, ответ на который сопоставим с созданием интеллекта.

В книге некоторые идеи даны лишь в виде набросков, сложные теории нейронных сетей и распознавания образов излагаются на уровне общих положений, но вы читаете не научную монографию. Я ставил перед собой цель ввести своего читателя в проблему, не требуя специальных знаний и хорошей математической подготовки, поэтому перед вами – лишь популярное изложение. Но я надеюсь, что, прочитав эту книгу, вы получите хорошее представление о том, чем занимаются специалисты по искусственному интеллекту, что это даст человеческой науке и технике и насколько до сих пор на самом деле неясен сам этот термин «интеллект».

Глава 1

Задача тысячелетия

Природа создала человека очень слабым. Мы медленнее всех бегаем, у нас нет острых зубов и когтей, мускульная сила оставляет желать лучшего, и даже способность наблюдать окружающий мир с помощью органов зрения и слуха сомнительна в сравнении с возможностями животных. Единственное наше преимущество, которое, впрочем, оказалось в эволюционной борьбе решающим, – это способность мыслить, делать выводы, накапливать знания. Разум дал возможность компенсировать все недостающее. Человек начал развивать науку в самом общем смысле этого слова, наука дала не только понимание окружающего мира, но и способы создания искусственной среды, комфортной для человеческого существования. Мы придумали технику, заменившую мускульную силу, появились устройства и технологии, не просто способные сдвинуть большой каменный блок или повалить дерево, но могущие выполнить тонкую работу с глубоко дифференцированными движениями. Машины стали делать работу, в принципе непосильную человеку, не по энергетике, а по сложности движений.

Все это было здорово, но интересно другое. В то время как люди уже сотни лет пользовались сложнейшими машинами в помощь своему телу, устройства, используемые в помощь интеллекту, были на удивление примитивны. Мы уже летали в небе своей планеты, строили железные дороги, высотные дома, в общем, много чего умели, и в это же время наши интеллектуальные помощники – это арифмометр, логарифмическая линейка и т. д., в общем, устройства, принципиально не отличающиеся от бухгалтерских счетов.

Наконец, во второй половине XX века произошло событие, качественно меняющее ход развития человеческой цивилизации. Наш разум впервые получил очень серьезного помощника в лице ЭВМ (элек-

тронно-вычислительной машины). И хотя прабабушка современного компьютера – ламповая ЭВМ – занимала большую комнату, было ясно, что это не просто большой арифмометр. Ее фотография представлена на рис. 1.1.

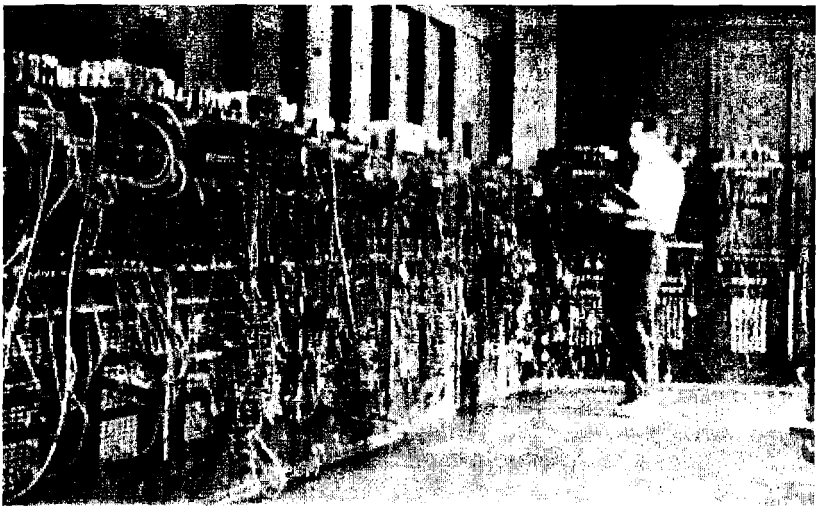

Рис. 1.1 ❖ «Эниак», ламповая ЭВМ

Даже первые машины могли перестраивать свою работу в соответствии с выполняемой программой. Правда, первые программы вводили в память машины не с клавиатуры, которой еще не было, но это уже были программы.

Надо сказать, однако, что первые разумозаменители, как и их потомки с терабайтами памяти и многоядерными процессорами, обладают принципиальным отличием от машин, заменяющих мускульное усилие.

Развитие техники медленно, но верно идет по пути исключения человека из выполняемого действия. Кое-где это уже произошло. Достигли своего идеала станки с ЧПУ, самолеты-беспилотники, системы управления атомными станциями и т. д. Человек здесь опустился (или поднялся, это кому как нравится) до функции чистого контроля. Компьютер же для своей работы потребовал новой касты высококвалифицированных специалистов, называемых программистами, которые и обеспечивают работу интеллектуальных машин, и это в то

время, когда мускульное усилие в производстве неотвратимо теряет свои позиции. Стало ясно, что машины могут заменить человека во всех видах активной, но не творческой деятельности. Большая же и важнейшая часть интеллектуальной работы для компьютера сегодня так же недоступна, как и для первых ЭВМ. Осмелюсь даже предположить, что и появление в некоем абстрактном будущем квантовых компьютеров с работоспособностью, на порядки превышающей все, что мы имеем сейчас, и практически безграничной памятью в этом смысле ничего не изменит.

И это «НО» означает, что появление компьютера – только преддверие к революции. Умение быстро обрабатывать большие объемы информации и выполнять неосильные для человека вычисления – это хорошо, но не так уж принципиально. А вот настоящая революция, способная изменить самое лицо нашей цивилизации, произойдет с появлением на планете искусственного интеллекта. Осталось только понять, что это такое, и затем реализовать его в осязаемом устройстве.

Интеллект – это, собственно, что такое?

Для создания искусственного разума надо понять, что такое интеллект вообще. Вопрос очень непрост. Например, если разум определить как способность разрешать проблемы теоретической физики, то окажется, что мало кто на планете обладает интеллектом. Определить же его как качественную характеристику именно человека означает ограничить понятие его человеческой разновидностью, и, кроме того, в таком определении не содержится никакой полезной информации, а значит, придется отвечать на вопрос: что же такого имеет человек, что свойственно только ему?

Как говорят мудрецы, чтобы вещью овладеть,
ей надо дать имя

Поэтому получить точное определение все же хотелось бы. Представьте себе, что вам принесли некую машину (или привели внеземную зверюшку), и утверждается, что она обладает интеллектом. Как это проверить? Нужны определение интеллекта (критерий) и точная процедура проверки этого критерия. Рассмотрим такое рабочее определение:

> Интеллект – качество психики, состоящее из способности адаптироваться к новым ситуациям, способности к обучению на основе опыта, пониманию и применению абстрактных концепций и использованию своих знаний для управления окружающей средой.

Оно опирается на понятия, которые, будучи интуитивно ясны, все же неточны и нуждаются в доопределении. К этому определению можно задать слишком много вопросов:

- Что такое знание?
- Что такое абстрактные концепции?
- Что значит обучение на основе опыта? Обучить дрессировкой можно очень многие виды животных. Означает ли это наличие у них интеллекта? А что такое обучение без дрессуры? В наших, человеческих школах очень многое делается именно дрессурой, означает ли это, что люди – школьники не обладают интеллектом?
- Что считать новой ситуацией и что считать успешной адаптацией? Очень часто люди теряются в необычной ситуации, означает ли это, что они не разумны?

Можно привести и другие, столь же мудреные и ненадежные определения. В чем их недостаток? Они для определения интеллекта используют сложные понятия. Это грубая логическая ошибка. Хорошее определение может свести понятие к более общему с добавлением достаточного уточняющего признака. Например, белая береза – это дерево с белой корой. При условии что понятие «дерево» уже определено и больше деревьев с белой корой не существует, это определение вполне удачно. Причем черную березу (а такая тоже есть) мы в определении отмели, так как ограничились именно белой.

Еще хорошее определение может свести термин к комбинации более простых. Например, окружность – это множество точек плоскости, равноудаленных от одной точки, называемой центром. Понятия точки, множества, плоскости проще, нежели понятие окружности. Кроме того, этим определением мы уходим от необходимости пояснять такие сложные вещи, как непрерывность и кривизна линии.

А еще хорошее определение может воспользоваться понятием эквивалентности. Например, синий цвет – это цвет неба. Круглый – это форма футбольного мяча. В общем, способов дать хорошее определение достаточно много. Заметим еще, что хорошее определение должно быть рабочим, то есть применимым для идентификации объектов. Это тоже не всегда просто. Например, можно определить километр как расстояние, которое способен пройти взрослый человек по прямой, не торопясь, за 10 минут. Допустим, что любой взрослый человек идет, не торопясь, с одной и той же скоростью (это не так, но допустим), то есть это точное определение. Но оно все равно не рабочее, так как измерять таким способом расстояния в сотни километров нереально.

Итак, какой способ нам выбрать? Свести к более общему с указанием определяющего свойства не получится. К сожалению или к счастью, но нам не известен ни один пример инопланетного разума. Животным, живущим рядом с нами, мы в разуме отказываем. Это означает, что у человечества нет возможности определить интеллект «вообще», без привязки к его человеческой разновидности. Разложить на более простые вещи не получится, для этого надо хорошо понимать, как он работает и из чего состоит. Этот метод фактически работает только на простых объектах. *Единственное, что нам остается, – это признать интеллект свойством, присущим каждому человеку (здоровому в медицинском смысле), и выделить некие вещи, которые может делать только человек, и только потому, что он разумен. Таким образом, мы для построения теста на интеллектуальность вынуждены выбрать человека как эталон.*

Критерий Тьюринга

Европейский математик Алан Тьюринг предложил следующий критерий:

Рис. 1.2 ❖ Алан Мэтисон Тьюринг

Для справки. Изображенный на рис. 1.2 Алан Мэтисон Тьюринг – английский математик, логик, криптограф, оказавший существенное влияние на развитие информатики. Кавалер Ордена Британской империи, член Лондонского королевского общества. Предложенная им в 1936 году абстрактная вычислительная «машина Тьюринга», которую можно считать моделью компьютера общего назначения, позволила формализовать понятие алгоритма и до сих пор используется во множестве теоретических и практических исследований.

«Человек взаимодействует с одним компьютером или одним человеком (с кем он говорит, ему неизвестно). На основании ответов на вопросы он должен определить, с кем он разговаривает: с человеком или компьютерной программой. Если человек не сможет сделать правильный выбор, значит, компьютерная программа обладает интеллектом».

Предложенный Тьюрингом критерий основан на том убеждении, что человеческая речь, имеется в виду, конечно же, осмысленная речь, есть неотъемлемая часть разума. Поэтому если я – разумное существо и мой собеседник может неограниченно со мной беседовать, и я не могу уличить его в бессмысленности, то он тоже разумен.

Критерий Тьюринга выглядит очень солидно, но необходимо сделать важное уточнение. Например, если я – специалист по переработке нефти, то я не имею права требовать от партнера знаний о процессе переработки. А это значит, что возможные бессмыслицы, которые он скажет, если побоится сознаться в своем незнании, не в счет. А утайка некомпетентности для разумного человека – дело, достаточно обычное. Это означает, что в тесте допустимы только темы общего порядка. Но и здесь мы пойдем по тонкому льду. Я знаю, что Волга впадает в Каспийское море. Мой партнер может этого не знать, следовательно, даже в бытовых темах нужно делать скидку на уровень образования. То есть незнание не эквивалентно бессмыслице.

Анализируя эти затруднения, легко прийти к выводу, что тест Тьюринга содержит в себе серьезное противоречие. С точки зрения теста, мы обязаны для обеспечения чистоты теста определить, что такое осмысленная речь. Фактически требуется понятие интеллекта определить через очень сложное понятие осмысленной речи. Это своего рода тавтология.

Единственный выход из создавшегося положения – отбросить все наукоемкие и философски солидные рассуждения, признать человеческую речь осмысленной по определению, на всякий случай отбросить все умные темы, ограничившись общими понятия-

ми, и поговорить с программой. Однако в этом пункте разработчиков искусственного разума ожидало самое главное разочарование. Оказывается, так называемая осмысленная речь достаточно легко алгоритмизируется. Первый виртуальный собеседник появился уже в 1966 году – программа Элиза, созданная Джозефом Вейзембаумом. Элиза моделировала речевое поведение, используя технику активного слушания, например используя фразы «Продолжайте, пожалуйста». Такие собеседники стали создаваться регулярно программистами и лингвистами, интересующимися проблематикой искусственного интеллекта, и с учетом того, что тест Тьюринга не имеет математически строгой формулировки и не вполне понятно, о какой ситуации можно сказать определенно «Программа тест прошла», то ценность самого теста таяла с появлением каждого нового виртуального собеседника.

И наконец, самая большая проблема. Человек обладает способностью осознавать то, что он делает. В отношении машины, отрабатывающей успешно тест Тьюринга, остается вопрос, а понимает ли машина, что она ведет беседу. В этой книге мы еще вернемся к этому вопросу, а пока заострите свое внимание на том факте, что обнаружение рефлексии на себя представляет задачу особой сложности.

Конечно, хотелось бы завершить этот текст на позитиве. Вроде того: «И наконец, они придумали идеальное определение и построили разумную машину, начав новую историю человечества». Но ничего подобного не произошло. Сегодня, что такое интеллект, точно так же не понятно, как и на заре создания теории ИИ, а значит, как и прежде, не ясно, как нам реагировать на появление еще одной умной машины. Мы уже знаем, что выигрыш машиной у шахматного мастера, даже гроссмейстера, не означает разума, точно так же мы не обнаруживаем интеллекта у программ, управляющих в режиме реального времени сложнейшими техническими системами. С каждой новой разработкой мы просто становимся свидетелями еще одного доказательства, что превосходство в еще одной частной сфере не означает возникновения интеллекта у машины. Проблема тысячелетия по-прежнему остается открытой.

Однако теория искусственного интеллекта существует, существуют и одноименные технические системы, используемые в реальных задачах. Искусственный интеллект, например, управляет системами защиты боевого корабля, позволяя исключить человеческий фактор

в условиях динамичного и сверхманевренного боя, в котором есть необходимость отбиваться сразу от большого количества нападений ракетами, самолетами, торпедами, движущимися очень быстро и способными маневрировать. Это реально работает. Так что такое искусственный интеллект?

В истории развития теории разума произошла следующая метаморфоза. Ученые, довольно основательно побившись головой о стенку философского осмысления понятия интеллект, обнаружили, что для решения конкретных задач нет необходимости во всеобщем, универсальном, самообучаемом и т. д. интеллекте. Это примерно то же самое, как нет необходимости моделировать кузнеца-человека, чтобы построить машину, штампующую из металла конкретные детали. Интеллектуальные процессы тоже можно моделировать и создавать конкретные программы под конкретные задачи. Таким образом, современная теория искусственного интеллекта перешла к созданию технических систем, работающих так, как будто они разумны, но только в рамках вполне определенной задачи.

Это направление человеческой мысли сейчас доминирует, но, конечно, задача разобраться в том, как работает интеллект вообще, осталась, и, может быть, и она когда-нибудь будет разрешена, но сейчас мы уже понимаем, насколько это сложно в действительности, и сколько уйдет на решение времени, совершенно не понятно.

История проблемы

К задаче определения искусственного интеллекта можно подойти с разных сторон. На мой личный взгляд, наиболее сложный способ заключается в том, чтобы попытаться дать определение без привязки к нашему человеческому мышлению и дать строгое, математически точное описание. Почему это сложно? Да потому, что надо дать исчерпывающее определение в терминах, не включающих информацию о человеческом разуме. Может быть, это было бы проще, если бы удалось поговорить с парой-тройкой представителей инопланетных цивилизаций. Но такой возможности нет. Другие подходы более реальны. Можно попытаться создать математическую теорию, описывающую именно наше, человеческое мышление, и создать устройство, работающее по этой теории, но не копируя человеческий мозг. И можно исследовать мозг в деталях и сделать его точную копию.

Нейрофизиологические исследования головного мозга дали огромную груду информации. Сегодня мы знаем, какие функции локали-

зованы в специфических отделах головного мозга, а какие размазаны по всему человеческому мыслительному аппарату. Выяснен главный системотехнический принцип. Оказывается, сверхвысокая эффективность работы мозга обеспечивается очень примитивными элементами – нейронами (нервными клетками). Каждый нейрон в отдельности практически ничего не умеет, но именно в этом ничегонеумении и заключается секрет успеха. Их примитивность на самом деле проявляется как универсальность. Действительно, если некий исполнитель способен выполнять только простую операцию и не желает вникать в общую постановку задачи, то такого исполнителя можно вставить в любую схему, большое количество элементов которой за счет специальной организации уже будет способно на многое. Нейронная идея даже породила целое направление в теории искусственного интеллекта – так называемые нейронные сети. И хотя пока самая сложная нейронная сеть не способна приблизиться по своим возможностям к человеческому мозгу, это направление считается очень перспективным.

Что же касается математически точной теории разума, то здесь положение дел очень туманно. Пока хорошая математика присутствует только в описании нейронных сетей и в теории эвристических алгоритмов, немного, конечно, но пока так. И это несмотря на то, что попытки математического осмысления проблемы имеют очень долгую историю. Здесь ожидались большие прорывы в силу ошибочного сведения мышления к одной из его форм – строгому логическому выводу. Впрочем, может быть, никакой особой ошибки в этом и нет, просто логика – такая вещь, которая наиболее просто формализуется и поддается исследованию, а пытаться пройти простым путем свойственно для нашего разума.

Формализация мышления

Отметим сразу, что единственное устройство, позволяющее моделировать искусственный интеллект, – это компьютер, работающий под управлением алгоритмов, представляющих собой последовательность команд, каждая из которых должна быть однозначно понимаема. Плюс к этому компьютер способен выполнять в одно и то же время только одну команду алгоритма. Существование параллельных вычислений в этом смысле мало что меняет. Возможность параллельных алгоритмов означает существование внутри алгоритма независимых частей, что-то вроде более простых алгоритмов. Эти технические ограничения мы обязаны иметь в виду при всех дальнейших рассуждениях. Системотехнические ограничения очерчивают жесткие гра-

ницы возможного. А надо сказать, что архитектура вычислительной системы – на самом деле главный фактор эффективности. Настолько важный, что суперкомпьютеры стали таковыми не столько за счет высокой скорости работы процессоров, сколько за счет усложнения конструкции. Но пока даже нейронные сети – это не всегда реальные технические устройства, а лишь модели на базе традиционных компьютеров. И, несмотря на то что современная наука, электроника и теория алгоритмов уже уверенно видят новые горизонты производительности, старая добрая архитектура фон Неймана является основным техническим решением.

Первую попытку формализации мышления следует признать за Аристотелем. Конечно, вряд ли древний грек формулировал задачу построения искусственного интеллекта. В античности такая задача не являлась актуальной хотя бы потому, что для древних Земля была населена массой различных мыслящих существ. Современное желание разобраться в вопросе разума, как мне кажется, произошло от осознания уникальности человеческого мышления. И логику силлогизмов, созданную Аристотелем, следует признать попыткой математически точного описания мыслительных процессов. И попыткой, достаточно успешной для того времени.

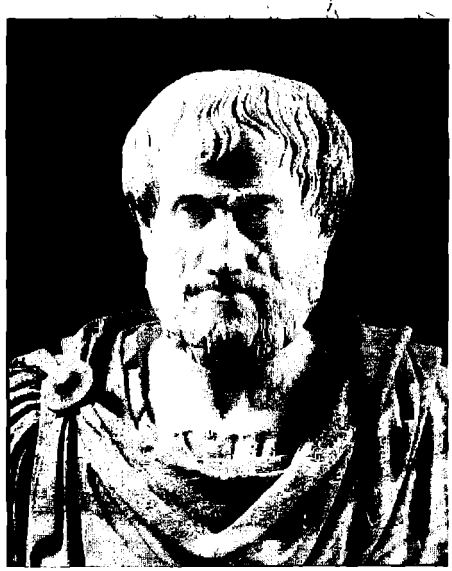

Рис. 1.3 ❖ Аристотель

Для справки. Аристотель – древнегреческий философ. Ученик Платона. С 343 г. до н. э. – воспитатель Александра Македонского. В 335 г. до н. э. основал Ликей (др.-греч. Λύκειο – Лицей), или перипатетическую школу. Натуралист классического периода. Наиболее влиятельный из философов древности; основоположник формальной логики. Создал понятийный аппарат, который до сих пор пронизывает философский лексикон и стиль научного мышления. Аристотель был первым мыслителем, создавшим всестороннюю систему философии, охватившую все сферы человеческого развития: социологию, философию, политику, логику, физику. Скульптура на рис. 1.3 естественно представляет предполагаемый облик философа.

Логика Аристотеля еще называется логикой силлогизмов. Силлогизм есть дедуктивное доказательство, состоящее из трех частей: большой посылки, меньшей посылки и заключения. Часто, иллюстрируя логику Аристотеля, в качестве примера приводят следующий силлогизм:

Все люди смертны (большая посылка).
Сократ – человек (меньшая посылка).
Следовательно, Сократ смертен (умозаключение).

Силлогистическая форма вывода истинных заключений присутствует во многих логических выкладках в более сложной форме. Мы не будем углубляться в развитие теории, заметим только следующее.

Интуитивно ясно, что силлогизмы дают слишком бедный аппарат для описания мышления. Чтобы в этом убедиться, достаточно посмотреть внимательно на доказательство *любой* сложной математической теоремы. Не так уж часто можно сделать вывод простым переходом от общего (Все люди смертны) к частному (Сократ смертен). В эту схему, например, не укладываются индуктивные рассуждения (переход от частного к общему). Совершенно не понятно, как логику силлогизмов использовать для доказательства простого утверждения «Ряд натуральных чисел бесконечен» или «$\sqrt{2}$ является иррациональным числом». Математика, по своей сути, не силлогистична. Еще менее полезна логика силлогизмов для естествоиспытателя, в чьей деятельности, например, есть такой внелогичный метод, как метод проб и ошибок, при использовании которого выдвигаются гипотезы, проверяемые экспериментом.

Логика силлогизмов не была успешна в *полной формализации* человеческого мышления, но начиная с Аристотеля стало ясно, что интеллект – не совсем целостная вещь. В нем есть различные компонен-

ты, например способность к логическому выводу. Для изучения этой способности начала развиваться наука – логика. Со временем логика смогла объяснить многое и выработать точные методы и принципы правильного логического мышления. Вот некоторые их них.

1. Из утвердительных (не путайте с истинными) суждений не может быть сделан отрицательный (не путайте с ложным) вывод.
2. Если одно из суждений отрицательно, то общий вывод будет отрицательным.
3. Закон тождества. Всякое суждение тождественно самому себе.
4. Закон непротиворечия. Два противоречащих суждения не могут быть одновременно истинными, одно из них обязательно ложно.
5. Закон исключенного третьего. Два противоречащих суждения не могут быть одновременно ложными. Одно из них необходимо истинно, другое ложно, третье исключено.
6. Закон достаточного основания. Всякая истина имеет достаточное основание.

О чем говорит, например, первый закон? Сколько бы у вас ни было о некоей вещи или ситуации утвердительных высказываний, ни одно из них не даст основания для отрицания. Например:

- Это яблоко красное.
- Это яблоко сладкое.
- Это яблоко кубанское.
- Это яблоко весит 200 граммов.

Ясно, что мы не можем на основании сказанного утверждать, что яблоко не круглое, но, может быть, можно утверждать, что оно не желтое? Не это ли означает красный цвет яблока, о чем утвердительно заявляется? С позиции обычной бытовой логики да, так, но с позиции строго формальной силлогизм тогда должен выглядеть следующим образом:

Это яблоко красное (большая посылка).
Красный цвет не есть желтый (малая посылка).
Следовательно, это яблоко не желтое (следствие).

Как видите, для необходимого отрицательного вывода в исходном наборе утверждений не хватает посылки «Красный цвет не есть желтый», а значит, построенный силлогизм не законен. Можно ли алгоритмизировать первый принцип из списка? Очевидно, да. Если в результате цепочки логического вывода получилось отрицательное суждение, то достаточно проверить набор исходных суждений, и если

среди них нет отрицательного, то логический вывод следует признать ошибочным. Единственно, заметим, что наличие отрицательных суждений в исходном наборе посылок еще не гарантирует истинности отрицательного заключения.

Разберем еще один закон. Последний, шестой – закон достаточного основания. Его можно интерпретировать следующим образом: *если некоторое суждение истинно*, то существуют набор истинных суждений и логическая цепочка, приводящая к искомому суждению от исходного набора. Этот закон также достаточно легко алгоритмизируется. Множество исходных суждений конечно. Следовательно, множество возможных логических цепочек (разумной конечной длины), которые можно построить на данном наборе суждений, так же конечно, а значит, достаточно построить все возможные логические цепочки и посмотреть, появится ли искомое суждение среди результатов. Если количество исходных суждений велико, то вычислительный процесс может занять время, столь длительное, что реально эта проверка окажется бессмысленной, но мы сейчас рассматриваем лишь теоретическую возможность, а вообще процесс когда-нибудь закончится, и мы получим вполне определенный результат.

Безусловно, современная формальная логика не исчерпывается шестью законами, это довольно развитая и сложная наука. Но раз нам удалось показать алгоритмизируемость двух законов, то можно надеяться, что алгоритмизация всей формальной логики – скорее дело большого труда, нежели принципа. Но вот беда – формальной логики для описания интеллектуальной деятельности явно недостаточно.

Во-первых, есть проблема целеполагания. Как поставить правильную цель, и что такое вообще правильная цель?

Рассмотрим простую ситуацию. Пусть процесс логического вывода имеет в своем начале только пять суждений. Для упрощения положим, что вывод осуществляется лишь в форме силлогизмов, и каждое исходное суждение может быть как малой, так и большой посылкой. Тогда имеем $2^5 = 32$ следствия. Теперь добавим эти следствия как возможные посылки к исходным и получим на втором шаге 2^{32+5} логических следствий. Это уже астрономическое число. Вывод неутешителен. Развивать любую науку во всех возможных и мыслимых направлениях невозможно. Процесс очень быстро потребует ресурсов, которых нет и никогда не будет у человечества.

Наша же наука способна развиваться и получать результаты за осмысленное время потому, что люди умеют ставить перед собой конкретную цель и определять направление исследований, продвигающее к поставленной цели. Ясно, что формальная логика не даст никаких средств для постановки цели, для выделения промежуточных целей, оценки полученного результата. Постановка цели – задача внелогическая, выполняемая какими-то другими механизмами, возможно, находящимися за пределами чистого мышления.

Во-вторых, не любая задача логически разрешима

Методы формальной логики ограничены в своем применении даже в очень простых задачах. Для иллюстрации рассмотрим ставшую уже классической проблему парикмахера. Эта задача достаточно сложно излагается в терминах теории множеств, но для ее популяризации придумана очень интересная и простая формулировка. Итак.

Рис. 1.4 ❖ Деревенский парикмахер

Условие (рис. 1.4 – внешний вид предполагаемого парикмахера). В некоей деревне живут мужчины. Женщины и дети там тоже живут, но нас интересуют только мужчины. Все мужчины делятся строго на две категории: мужчины, которые бреются сами, и мужчины, которых бреет парикмахер, других видов мужчин нет. Парикмахер – тоже мужчина, он тоже живет в этой деревне, и он один. Вопрос: кто бреет парикмахера?

Из условия задачи ясно, что есть возможность применить закон исключенного третьего. Действительно, для парикмахера есть только две возможности: либо он бреется сам, либо он не бреется сам. Это взаимоисключающие суждения, поэтому с необходимостью одно из них ложно, а другое истинно, третьего не дано. Так нам говорит закон исключенного третьего. Однако проведем рассуждения:

Суждение первое. Парикмахер бреется сам.

В этом случае парикмахер – это мужчина, который бреется сам, а таких мужчин не бреет парикмахер, а так как он и есть парикмахер, то, следовательно, он сам себя брить не может, следовательно, это суждение ложно.

Суждение второе. Парикмахер не бреется сам.

В этом случае парикмахер – это мужчина, которого бреет кто-то другой, но это означает, что его бреет парикмахер, а так как он и есть парикмахер, то получается, что он бреется сам, и мы получили противоречие. Следовательно, и это суждение ложно.

Итак, два суждения взаимоисключающие, оба ложны, а третьего не дано. Как быть в такой ситуации, формальная логика ничего сказать не может. Теоретически проблема решена Давидом Гильбертом, но способом, который просто ограничивает сферу деятельности формальной логики, а значит, решает задачу, убивая окончательно наши надежды положить формальную логику в основу искусственного интеллекта.

Вообще, вопрос, что делать с задачей, которая не решаема, – возможно, один из самых интересных в истории и философии человеческой науки. В этом отношении очень показателен разговоров двух героев братьев Стругацких из романа «Понедельник начинается в субботу» – двух магов: Федора Симеоновича Киврина и Кристобаля Хозевича Хунты:

– Г-голубчики, – сказал Федор Симеонович озадаченно, разобравшись в почерках. – Это же проблема Бен Б-бецалеля. К-калиостро же доказал, что она н-не имеет решения.

– Мы сами знаем, что она не имеет решения, – сказал Хунта, немедленно ощетиниваясь. – Мы хотим знать, как ее решать.

– К-как-то ты странно рассуждаешь, К-кристо... К-как же искать решение, к-когда его нет? Б-бессмыслица какая-то...

– Извини, Теодор, но это ты очень странно рассуждаешь. Бессмыслица – искать решение, если оно и так есть. Речь идет о том, как

поступать с задачей, которая решения не имеет. Это глубоко принципиальный вопрос, который, как я вижу, тебе, прикладнику, к сожалению, не доступен. По-видимому, я напрасно начал с тобой беседовать на эту тему.

И на самом деле это глубоко принципиальный вопрос. Хочу заметить, что самые большие открытия человеческая наука совершала, перескакивая через нерешаемые и не понимаемые здравым смыслом задачи. Пример тому – борьба с аксиомой параллельных. Есть два противоречащих суждения: параллельные прямые существуют, и параллельные прямые не существуют, – и это та самая ситуация, когда взаимоисключающие утверждения могут быть истинными. Евклид положил, что да, через точку, не принадлежащую данной прямой, можно провести одну и только одну прямую, не пересекающуюся с данной прямой.

Это утверждение с точки зрения Евклида является аксиомой, но уж больно по своей сложности оно похоже на теорему. Поэтому люди две тысячи лет пытались его либо доказать, либо опровергнуть. В XIX веке трое ученых: Гаусс, Лобачевский и Риман – догадались отбросить логические законы и положить, что любое суждение о параллельных истинно, если на его базе можно развить геометрию. Так появились неэвклидовы геометрии и совершенно новое понимание свойств пространства и заодно ограниченности формальной логики.

Вернемся к критике формальной логики. Еще древние обнаружили существование парадоксов. Парадокс – это ситуация, когда вроде бы посылки для логического вывода безупречны, сам вывод проведен строго, в полном соответствии с законами логики, но полученный результат откровенно ложен, до нелепости ложен. Одним из первых логиков, описавших такие ситуации, был древнегреческий философ Зенон. Его умозаключения называются апории Зенона. Приведем для примера один из них.

Ахиллес и черепаха

Ахиллес – это древнегреческий воин, могучий, как все мифологические герои. Соответственно, он и бегает быстро. Что такое черепаха, думаю, объяснять нет необходимости. И вот эти двое решили, по Зенону, побегать наперегонки. Ахиллес, понимая, что черепаха бегает несколько медленнее, дал ей фору. То есть сначала стартует черепаха, и лишь спустя некоторое время Ахиллес. А теперь, как говорят фокусники, следите за руками (рис. 1.5 – иллюстрация к задаче).

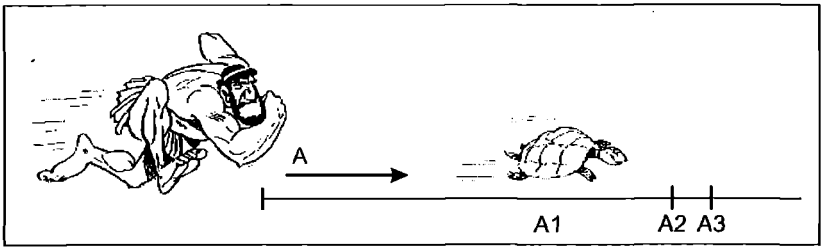

Рис. 1.5 ❖ Ахиллес и черепаха

В начале старта Ахиллеса между ним и черепахой есть некоторое расстояние. На его преодоление Ахиллесу нужно некоторое время. Пусть, например, 10 минут. Через 10 минут Ахиллес прибудет в точку, в которой была черепаха, но ее там уже нет. За эти десять минут черепаха пройдет какое-то расстояние. Преодолеть новую дистанцию Ахиллесу труда не составит, но на это опять уйдет время. За это время черепаха еще что-то там пробежит. И получается, что как бы Ахиллес не старался, между ним и черепахой всегда будет какое-то расстояние, для преодоления которого Ахиллесу нужно время, но черепаха за это время пройдет новое расстояние, *а значит, между Ахиллесом и черепахой всегда будет непройденное расстояние, а значит, Ахиллес никогда не догонит черепаху!*

Решение проблемы лежит в области теории бесконечно малых. Сегодня эта теория называется дифференциальным исчислением. Во времена Зенона такой теории не было, а в рамках формально-логических систем проблема не разрешима. Это хороший пример ограниченности формальной логики и отличия интеллекта от его частного инструмента – логического вывода. Все сказанное здесь являет нам печальную истину – все достижения человеческой логики в области формализации мышления, скорее, показали ограниченность логики, чем ее силу. Проблема интеллекта оказалась неизмеримо сложнее.

Психология мышления

Если разум не удалось объяснить с позиции формальной логики, то это не означает принципиальной необъяснимости предмета. Просто не с того конца подходили к вопросу. Изначально философам надо было бы заметить, что даже очень неумный человек, без какого-либо образования, не имеющий никакого представления о логике, должен

быть признан нами разумным. О чем это говорит? Да о том, что интеллект – явление внелогического порядка. Его природа другая. Есть смысл признать интеллект психологическим явлением и перейти в новую сферу – сферу психологии. Посмотрим, что удалось добиться в деле объяснения разума психологам.

Например, определение предмета психологии мышления П. Я. Гальперины звучит так: «Психология изучает не просто мышление и не все мышление, а только процесс ориентировки субъекта при решении интеллектуальных задач на мышление». Таким образом, с точки зрения одного из лучших советских психологов эта наука не претендует на полное решение задачи исследования интеллекта, а желает лишь решить вспомогательную проблему.

Другой столп советской психологии А. Н. Леонтьев определяет мышление как высшую ступень познания. Звучит тоже не слишком обнадеживающе. С таким же успехом можно мышление определить как форму разума, разум – как форму интеллекта, а интеллект – как способность к мышлению (то есть пойти по кругу из тавтологий). Может быть, этот сарказм и излишен, советские психологи сделали довольно много для понимания сути механизмов мышления, но я хочу проиллюстрировать мысль – психология мышления не решила задачу определения интеллекта и не создала точных теорий.

Может быть, в этом вопросе немного дальше продвинулись психологи западной науки?

Психология относится к тем наукам, которые объясняют человека и общество, поэтому психология всегда была сильно подвержена идеологическим влияниям. Можно предложить, что советская идеологическая установка настолько сильно повлияла на науку, что не позволила ей прийти к решению, которое было где-то рядом, но тогда нужно обратиться к западной психологии.

Жан Пиаже определяет мышление как способность психической адаптации к новым условиям. Интеллектуальный акт – это «акт внезапного понимания». Согласитесь, как-то совсем не конкретно. А гештальтпсихология основную идею, которую начал разрабатывать Вертгеймер, взяла за основу утверждение, что акт психического осознания не разлагается на составные части и может быть исследован только как целое. Но любая алгоритмизация потребует аналитики, выделения составных компонентов, отдельных процессов, приводящих к мыслительным результатам. Некоторое время была весьма популярна теория ассоциативного мышления. Вот она, пожалуй, из тех

теорий, которые взялись за труд выявить конкретные мыслительные механизмы.

Ассоциация – это связь между отдельными фактами, событиями, предметами или явлениями, отраженными в сознании человека и закрепленными в его памяти. Ассоциативное восприятие и мышление человека приводят к тому, что появление одного элемента, в определенных условиях, вызывает образ другого, связанного с ним.

По мнению основателя ассоциативной психологии, английского врача Д. Хартли (1705–1757), ассоциативное мышление – понятие, отражающее факт использования в мышлении закона ассоциации (сочетания): любая связь представлений и действий выводима из ощущений и оставленных ими следов в мозгу. Например, ученик, решавший задачу некоторое время назад, помнит логическую цепочку, приведшую к составлению квадратного уравнения. Получив новую задачу, с похожим условием, он включает ассоциативную связь и пробует пройти той же дорогой. Если условия двух задач действительно похожи, то нет ничего невероятного, что этот путь опять приведет к квадратному уравнению.

Понятие ассоциации в психологии разработано достаточно хорошо и в плане определения, и в плане описания механизмов работы ассоциативного мышления, настолько хорошо, что сомневаться в реальности существования такого типа мышления уже не приходится. Но чем точнее и полнее мы сможем описать ассоциативное мышление, тем точнее и полнее встанет и другая правда, что это всего лишь один из механизмов, некий частный случай, не решающий задачи в целом. Ассоциации, например, не объясняют нашу способность к обобщению, не объясняют существования абстракций, процесс формирования цели и многое другое.

В общем, надо признать, что психологи, логики, математики, кибернетики сделали очень много для понимания частных механизмов мышления, но чем теории становились детальнее, тем отчетливее проступал факт нерешаемости вопроса в целом. Можно описать интеллект как деятельность сознания, формализовать понятие гештальта, алгоритмизировать способность к аналитике, синтезу в рамках той или иной формальной схемы, но ответ на главные вопросы все равно ускользает:

- Каким образом интеллектуальная система способна самообучаться без ограничения областей знания?

- Что такое знание, как оно используется для получения нового знания?
- Как феномен интеллекта связан с феноменом сознания?
- Что означает создать искусственный интеллект?

Эвристические алгоритмы

Заход на проблему со стороны формальной логики, психологии, вместе с попыткой увязать наметившееся понимание с возможностями имеющейся цифровой техники, высветил очень серьезную проблему – большой разрыв между сложностью задачи и имеющимися ресурсами моделирования. Этот разрыв принципиален. Компьютерный алгоритм в классическом понимании (здесь надо оговориться, я имею в виду понимание, существовавшее на заре развития компьютерной техники) – вещь, железно приводящая к одному и тому же результату вне зависимости от количества запусков алгоритма. Миллион раз запускаем, миллион раз получаем один и тот же ответ при одних и тех же входных параметрах. А если входных данных не хватает, то алгоритм просто не работает.

Интеллектуальная система, напротив, может начать работу и при недостатке данных, и даже острая нехватка информации не становится препятствием для получения результата, пусть и не всегда удовлетворительного. Интеллектуальная система не работает в строгих рамках. Она способна выбирать путь из нескольких вероятных. В общем, она способна работать эвристически.

> *Эвристика – это основанное на опыте правило, существенно ограничивающее поиск решения в сложной задаче. Эвристика не гарантирует оптимальности полученного решения, полезная эвристика предлагает варианты, которые с высокой долей вероятности оказываются достаточно хорошими.*

Прежде чем двигаться дальше, позвольте привести простой пример эвристического алгоритма. В учебниках по программированию можно встретить задачу о двух кучах камней. Ее условие таково: есть одна большая куча камней, возможно разного веса. Требуется раскидать ее на две кучи так, чтобы между ними была минимальная разница в весе.

Вообще, если нам не нужен реальный результат за ограниченное время, то задача решается очень легко. Загоним камни исходной кучи в массив и построим из полученного таким образом массива все возможные сочетания камней. Каждое сочетание – это одна куча,

а оставшиеся вне сочетания камни – другая. Для каждой полученной таким образом пары определим разницу в весе и выберем из всех пар ту, для которой разница минимальна.

Проблема в том, что этот алгоритм переборный. А количество всех возможных сочетаний из N элементов равно 2^N. То есть даже при очень небольшой исходной куче, например в 100 камней, общее количество сочетаний 2^{100}. И получается так, что решение есть и его как бы нет. Дождаться, когда компьютер его обнаружит, человеческой жизни не хватит. А теперь давайте откажемся от желания найти идеальное решение. Пусть нам будет достаточно решения хорошего. Тогда возможен такой алгоритм:

- Упорядочим исходную кучу камней в порядке убывания веса.
- Пока в исходной куче камней есть хотя бы один камень, делаем:
  - берем очередной камень;
  - если правая куча тяжелее левой, то кладем очередной камень в левую кучу, иначе кладем его в правую.

Проиллюстрируем алгоритм примером. Пусть исходная куча содержит такие камни: (9, 15, 1, 1, 7, 4).

После упорядочивания массив примет такой вид: 15, 9, 7, 4, 1, 1.

Шаг 1: Правая – 15; Левая – 0; Исходная 9, 7, 4, 1, 1.
Шаг 2: Правая – 15; Левая – 9; Исходная 7, 4, 1, 1.
Шаг 3: Правая – 15; Левая – 9 + 7 = 16; Исходная 4, 1, 1.
Шаг 4: Правая – 15 + 4 = 19; Левая – 9 + 7 = 16; Исходная 1, 1.
Шаг 5: Правая – 15 + 4 = 19; Левая – 9 + 7 + 1 = 17; Исходная 1.
Шаг 6: Правая – 15 + 4 = 19; Левая – 9 + 7 + 1 + 1 = 18; Исходная – Пусто.

Заметим, что за шесть шагов было найдено идеальное решение. А алгоритм полного перебора отработал бы $2^6 = 64$ варианта. Вот что такое эвристический алгоритм. Его суть в том, что принимается допущение, которое не обязательно верно во всех случаях жизни, но выглядит вполне правдоподобно. Например, рыбак, выбирая место для ловли, может рассуждать так: «Вон там я вижу омут. В омуте может водиться крупная рыба. Попробую я, однако, порыбачить там». Конечно, рыбак может и ошибаться. В этой реке вообще может рыбы не быть. Но предположение, что в омуте она есть, выглядит разумно, и это эвристическое допущение.

Эвристика создает качественно новые возможности для разработки эффективных алгоритмов, вводя в наш инструментарий такую вещь, как опыт

Эвристика все меняет радикально. Классический алгоритм принимает во внимание только входные данные, и если они такие же, как и в предыдущем запуске, то и ответ будет тем же. Эвристический алгоритм, кроме входных данных, имеет возможность оценивать опыт. В примере с рыбаком это работает так: «Я имею перед собой реку, в ней есть отмель и есть омут (это примерно так, как выглядит на рис. 1.6). Я уже десять раз встречал такую ситуацию и восемь раз из десяти был успешен, порыбачив в омуте. Значит, есть смысл попробовать еще раз».

Рис. 1.6 ❖ Эвристический рыбак

Заметим, что для нашего рыбака первая река с омутом и отмелью не несет в себе никакой дополнительной информации. Но уже первый заход создает новую информацию, называемую опытом. И что любопытно, эту информацию рыбак может использовать и на другой реке.

Опыт дает возможность закреплять успешную стратегию. Две удачные попытки из трех подвигнут рыбака попробовать обнаруженную стратегию еще раз, но на треть останется сомнение в ее эффективности. Если попытка опять будет удачной, то на следующий раз доля сомнения уменьшится до четвертой. Это явление называется закреплением успешной стратегии.

Эвристика позволяет менять стратегию. Рассмотрим это опять на примере с рыбаком. Допустим, у него поменялись интересы к видам рыб, и он перешел на рыбу, которая больше любит отмели, но рыбак пока этого не знает. Первые попытки ловли другой рыбы естественно пытаться реализовать в рамках отработанной стратегии омута. Тем более что стратегия основательно закреплена и доля сомнения очень низка. Но каждая неудача ловли на омуте будет увеличивать уровень сомнения. Введем понятие порога успешности. Назовем таким порогом величину сомнения, превышение которой означает признание провала стратегии. Пусть пороговым значением для рыбака будет половина неудачных попыток в их общем числе. Тогда если он закрепил стратегию 9 успешными попытками (на прежней рыбе) из десяти, то 8 неуспешных приведут его к пониманию непригодности старой стратегии в новых условиях.

А мы можем методику определения неуспешности эвристики усилить. Пусть, например, замечено, что ранее из трех попыток рыбачить в омуте две были железно успешными. Тогда три неудачи подряд (а не 8) могут быть поводом для переосмысления стратегии.

Эвристика может быть разноуровневой. Например, опыт рыбалки в омуте можно не распространять на рыбу вообще. Можно ввести более общее правило, требующее пересматривать стратегию каждый раз при переходе на новый вид рыбы, или новую снасть, или новую наживку.

Если вы понаблюдаете за собой, то придет понимание, что эвристика есть основа принятия практически любого решения. Почти всегда в бытовых ситуациях в наших суждениях лежат недоказанные, но внешне разумные допущения: в большом магазине проще найти нужный товар; чтобы устроиться на хорошую работу, необходимо быть прилично одетым; дорогой товар более качественен. На любое из этих или других допущений можно возразить, но тем не менее мы пользуемся сотнями разных эвристик, даже не задумываясь об их доказательстве. Люди справедливо полагают наличие опыта достаточным основанием.

Еще одна интересная возможность, открытая для эвристических алгоритмов, – это передача знания. Эвристическая программа может использовать не только свой опыт, но и опыт другой эвристической программы. Конечно, для этого необходимо решить очень много сложных вопросов. Например, необходимо дать практически полез-

ное определение знания, придумать форму его хранения, процедуру передачи. Необходимо придумать язык для записи знаний, и это не будет классический язык программирования.

Дополнительно заметим, что задача определения понятия «знание» не равнозначна задаче определения понятия «интеллект». Нас интересует определение, полезное для эвристической машины, именно полезное, а не всеобъемлющее. А это очень большая разница. Например, для описания движения планет вокруг Солнца нет необходимости в общей теории относительности, можно даже обойтись без Всемирного закона тяготения Ньютона. Вполне достаточно законов Кеплера. Для того чтобы построить паровую машину, нет необходимости в исчерпывающем понимании сущности физической энергии. Вполне достаточно понимать, как энергия пара преобразуется в энергию механическую. Так же и нам достаточно определить вид знания, полезного для эвристической машины. Впрочем, и этот вопрос достаточно сложен, мы его обязательно обсудим в другой главе, а пока ограничимся примерами рыбака и рыбки.

В заключение

В своем дальнейшем развитии теория искусственного интеллекта пошла вполне разумным путем – разделив свой предмет на два не очень связанных между собой вопроса. Первый вопрос глобальный, вселенского значения: что такое интеллект и как создать систему, как минимум разнозначную человеку, а может быть, его и превосходящую. Второй вопрос: как для каждого отдельно взятого интеллектуального процесса создать его эффективную техническую модель.

Большая часть этой книги, ее последующих глав будет посвящена именно этой второй, реально решаемой задаче, но некоторое время мы потратим и на первую. В заключение главы хотелось бы обратить ваше внимание еще на один интересный момент. Сейчас в научной и особенно в околонаучной прессе все чаще появляются разного рода предсказания даты создания искусственного интеллекта. К этим суждениям и профессионалов, и дилетантов надо подходить с большой осторожностью. Если говорить о большой задаче, то предсказывать дату ее полного решения не очень серьезно. Пока не вполне ясно, что она из себя представляет. Человек еще обладает такими вещами, как сознание, воля, и не факт, что эти составляющие сводятся к интеллек-

ту, и совершенно не факт, что существо, идентичное человеку, определяется именно интеллектом. В общем, здесь еще надо разбираться и разбираться, о чем идет речь и чего можно добиться.

Что же касается прикладного понимания искусственного интеллекта, то он уже давно живет рядом с нами. Днем его рождения, например, можно считать день первой шахматной партии, выигранной машиной у гроссмейстера. Технические системы с элементами искусственного интеллекта входят в нашу жизнь незаметно, но очень уверенно, они уже накопили потенциал такого масштаба, что, наблюдая их работу, можно подумать и о реальности иного разума, пришедшего не из глубин космоса или мирового океана, а созданного нашими же усилиями.

Глава 2

Вся жизнь – игра

Есть один вид человеческой деятельности в высшей степени интеллектуальный, но тем не менее, как оказалось, легко поддающийся алгоритмизации. Я имею в виду игру. Первые программы, эффективно играющие против человека, появились еще на заре развития вычислительной техники. Что интересно, штурм игры искусственным интеллектом пришелся на шахматы, каковые следует признать одной из наиболее сложных игр. И успех пришел достаточно быстро. Начало шахматных разработок можно отсчитывать от статьи Клода Шеннона, опубликованной еще в 1949 году, в которой рассмотрены основные проблемы игровых моделей. Шеннон впервые представил игру в виде дерева позиций, ветви которого – это возможные ходы, а выбор хода – это выбор наилучшего продолжения из возможных. Процедура выбора получила название минимаксной. Ее название характеризует самую суть действия машины, и немного позже мы это обсудим.

Рис. 2.1 ❖ Клод Элвуд Шеннон

Клод Элвуд Шеннон (рис. 2.1) – американский инженер и математик, его работы являются синтезом математических идей с конкретным анализом чрезвычайно сложных проблем их технической реализации. Является основателем теории информации, нашедшей применение в современных высокотехнологичных системах связи. Шеннон внес огромный вклад в теорию вероятностных схем, теорию автоматов и теорию систем управления – области наук, входящие в понятие «кибернетика».

Первым проект Шеннона реализовал Тьюринг, но его программа, показав принципиальную приемлемость, тем не менее не смогла выиграть даже у слабого шахматиста. Первой если не сильной, то вполне играющей шахматной программой можно считать проект MANIAC-I, реализованный в 1956 году группой сотрудников Лос-Аламосской лаборатории. Эта программа наиболее точно соответствовала идее Шеннона, ей также не удалось показать выдающихся результатов, но одну из трех партий против слабого игрока она уже могла выиграть.

Дальнейшее развитие игровых программ было связано с двумя факторами: развитием математического аппарата игровых программ и ростом ресурсов вычислительных машин. На годы главной игрой, на которой упражнялись специалисты по искусственному интеллекту, остались шахматы, но со временем, с ростом доступности компьютеров, появлением большого числа программистов, руки стали доходить и до других игр. Сегодня в сети можно найти программы, играющие против человека в самые различные игры. Что же касается шахмат, то здесь были достигнуты просто великолепные результаты. Программы, использующие ресурсы персональных компьютеров, вполне успешно играют против хороших шахматистов, а гроссмейстеры-суперкомпьютеры смогли разделить шахматный пьедестал с гроссмейстерами-людьми.

В 1996 году состоялся первый матч Deep Blue с Гарри Каспаровым, в котором чемпион мира одержал победу со счетом 4:2. Deep Blue – это 6-процессорный суперкомпьютер, способный просчитывать 100 млн позиций в секунду. Через год состоялся матч-реванш с модернизированным 8-процессорным Deep Blue, считающим вдвое быстрее. Компьютер впервые победил лучшего шахматиста планеты со счетом 3,5:2,5. Пока компьютер не умел оценивать позицию, как человек, рост силы игры достигался по большей части за счет увеличения мощности «железа». Даже в качестве алгоритма перебора все еще использовался «брутфорс» (грубая сила), перебиралось как можно больше вариантов, но очень быстро. В 2003 году состоялся еще

один матч Каспарова против компьютера – Deep Junior, работавшего на 4-процессорной системе с процессорами Pentium IV 1,9 ГГц и 3 Гб оперативной памяти. Junior стал первой программой, демонстрирующей «человечную» игру и способной пойти на жертву ради инициативы. Матч закончился вничью.

Компьютер против человека. Как это выглядит в принципе?

Для меня лично, когда я впервые увлекся проблемами искусственного разума, было самым большим открытием то, что моделирование интеллекта выполняется на базе очень простых принципов. Их техническая реализация может быть трудоемкой и даже очень трудоемкой, так как есть необходимость учитывать огромное количество деталей, и одной формулой процесс не опишешь. Но базовых идей очень немного, и их понимание не требует исключительной одаренности.

В этой главе мы рассмотрим принципы создания программ, играющих в интеллектуальные игры. Информации главы вполне достаточно, чтобы написать несложную программу, которая будет вполне прилично играть, при условии что вы – технически грамотный программист. Конечно, я не советую пытаться соревноваться с коллективами, пишущими программы, играющие в шахматы или го. Для начала можно попробовать реализовать игру попроще.

А теперь поговорим о том, как это работает. Базовые идеи разберем на шахматах, думаю, что хотя бы правила этой игры известны всем. Две вещи сомнения не вызывают: для принятия решения об очередном ходе необходимо уметь оценивать ситуацию, насколько она хороша или насколько она плоха, и необходимо уметь просчитывать течение игры на некоторое количество ходов вперед. Просчет игры заключается в построении всех возможных вариантов с точки зрения «грубой силы» или всех разумных вариантов, если есть хорошая эвристика.

Как эту работу выполняет человек

В решении задач искусственного интеллекта возможны два общих подхода. Можно попытаться реализовать человеческий способ мышления, и можно придумать метод, решающий ту же задачу, но иначе, не по-человечески. Примеров успешной реализации второго

подхода масса. Например, движущийся по земле аппарат человек снабдил колесами, а не ногами. Наши летательные аппараты также принципиально отличаются от устройства птицы или летающего насекомого. А значит, и программа, играющая в шахматы или шашки, не обязана выглядеть по образцу и подобию своего создателя. Но тем не менее рассмотреть процесс игрового мышления человека полезно. На картинке (рис. 2.2) этюд мастера шахматной композиции Рихарда Рети.

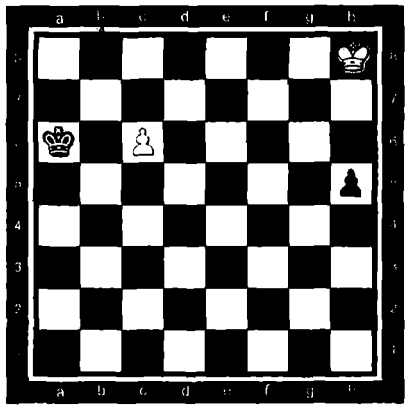

Рис. 2.2 ❖ *Этюд мастера Рети*

По замыслу автора, белые, начиная, должны свести игру к ничьей. Попробуем провести человеческий анализ этой позиции. В чем здесь проблема? Позиция содержит две угрозы. С одной стороны, черная пешка рвется к последней горизонтали, и белый король ее уже не догонит. Аналогичная угроза есть и для черных, белая пешка через два хода может стать фигурой, но черный король успевает ее перехватить, а белый не успевает прийти на помощь своей пешке. То есть перед белым королем стоят две задачи: помощь своей пешке и перехват черной – не решаемые по отдельности.

Однако если две задачи не решаются каждая в отдельности, это не означает, что не решаема и их комбинация. Белый король имеет возможность, двигаясь по черной диагонали, одновременно приближаться к зоне перехвата черной пешки и к зоне поддержки своей. Если черные будут выполнять движение пешкой, то белый король успеет

поддержать свою, а если черные решат уничтожить белую пешку, то белый король успеет перехватить черную. Таким образом, правильным решением для белых оказывается движение по диагонали.

Даже на этом, довольно простом примере видно, что игрок – человек использует довольно сложный понятийный аппарат: перехват, зоны гарантированной поддержки, угроза, одновременная угроза. Вопрос здесь вот какой: а имеем ли мы техническое устройство, способное оперировать понятиями такого смыслового уровня? Если взглянуть на проблему глазами специалиста середины XX века, то ответ отрицательный. Современный уровень развития систем искусственного интеллекта, конечно, другой, и сейчас движение в сторону моделирования человеческого способа мышления уже не выглядит таким уж невероятным. Но дело в том, что в интеллектуальных играх техника повторила прецедент колеса. В то время когда интерес к моделированию шахмат был очень велик, ресурсы для реализации человекоподобного шахматиста можно сказать что отсутствовали, и теория искусственного интеллекта нашла возможность решить поставленную задачу совсем иным, но очень эффективным способом. Коротко этот способ можно обозначить двумя словосочетаниями: «дерево перебора» и «оценочная функция».

Первая базовая идея – дерево перебора

Допустим, шахматная ситуация допускает десять ходов игрока. Конечно, реально в шахматной партии, даже если осталась пара фигур, вариантов хода существенно больше, чем десять, но мы для упрощения анализа остановимся на числе в 10 продолжений. Тогда его противник тоже имеет выбор из 10 ходов. Таким образом, ход игрока и реакция его оппонента создадут 100 позиций. Еще пара ходов (игрок – противник), и конечных позиций уже 10 000. Простая арифметика дает астрономическое число. N ходов обоих игроков породят 10^N ситуаций. Таким образом, партия в 100 ходов даст 10^{100} конечных позиций. А значит, даже если бы в каждой позиции действительно было бы возможно только 10 ходов, количество конечных ситуаций практически необозримо.

Есть, правда, вопрос: а зачем анализировать возможные продолжения от исходной позиции до конечной? Вопрос вполне правомерный. Люди-шахматисты, очевидно, не выполняют подобной работы, однако партии в исполнении людей выглядят вполне разумно. Есть, кстати, любопытная легенда об одном из величайших шахматистов

всех времен и народов – Хозе Рауле Капабланке. Вроде бы его как-то спросили, как далеко он продумывает игру, на что Рауль ответил: на один ход. Значит, это возможно – не смотреть глубоко.

Вспомним, однако, что человек – довольно сложно устроенная машина. У нас такой нет. Мы не можем опираться на опыт, нет и возможности использовать развитую шахматную теорию, что, конечно, минус. Знание теории дает довольно много. Допустим, требуется выяснить, выдержит ли некая конструкция удар кувалдой. Знание теории сопротивления материалов позволяет провести расчет и дать ответ без кувалды. Если теория испытателю неизвестна, то остается только выполнить удар. Примерно такова же ситуация и в шахматах. Если мы знаем, что потеря центра в дебюте – это почти всегда плохо, и если мы видим, что такая угроза существует, то нет необходимости анализировать все продолжения, будет вполне достаточно сосредоточиться на угрозах центру. Если такой теоретический факт неизвестен, то *придется провести анализ на некоторое количество ходов вперед*, чтобы обнаружить то, о чем теория говорит сразу. Конечно, шахматная теория существует, но давайте пока осложним себе жизнь и положим, что она нам неизвестна. Итак, зафиксируйте свое внимание на проблеме. Обход дерева перебора необходим, но полный его обход *невозможен*. Продолжим анализ.

Чем меньше мы знаем,
тем больше возможностей необходимо проверять

А значит, есть жесткая необходимость уметь строить дерево перебора от текущей позиции вглубь на как можно большее число ходов. Это фундаментальная необходимость. Конечно, повторимся, для шахмат существует глубокая и хорошо разработанная теория. Придумать форму записи этой теории в виде, *пригодном для компьютерной программы*, сложно, но можно. Но мы ведь говорим об игре вообще. А игр человечество придумало сотни, и для подавляющего их большинства никакой теории нет вообще. Перебор же в силу своей простоты реализации возможен всегда. Есть и интересная особенность игрового перебора. *Все возможные ходы представляют собой* не хаотичное множество, а довольно строгую структуру дерева. Что-то вроде такого, как на картинке ниже (рис. 2.3).

При наличии такой структуры все, что мы должны сделать как программисты, – это научиться обходить дерево в поисках ветки, обладающей нужными свойствами. А для того, чтобы уяснить, что такое «нужные свойства», рассмотрим вторую базовую идею.

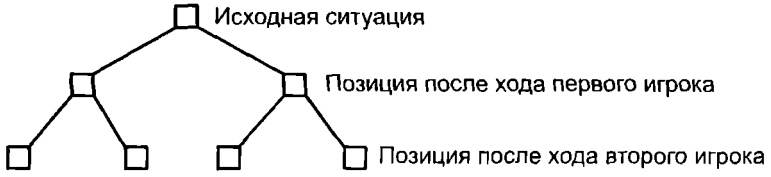

Рис. 2.3 ❖ Дерево перебора

Вторая базовая идея – оценочная функция

Итак, допустим, мы в анализе позиции продвинулись по дереву перебора на несколько ходов вперед и получили некоторое количество позиций, которые будем считать конечными. Необходимо их оценить. Самая примитивная ситуация – это кто-то уже выиграл. Возможно, будут и такие ситуации, но, скорее всего, каждая из конечных позиций находится в подвешенном состоянии: нет очевидного проигрыша, нет и очевидной победы, но что-то изменилось, и это что-то требуется измерить простыми количественными показателями. Иначе говоря, необходима числовая оценка, характеризующая качество позиции.

Заметим сразу, что оное качество определяется разными видами факторов. Очень важно, какой материал находится на доске. Материальное преимущество, в принципе, может быть решающим. Например, если в игре в шашки у вас четыре дамки, а у противника одна шашка и ход за вами, то совершенно не важно, как расположены шашки. Если у вас на доске к концу шахматной партии остались ферзь и король, а у противника только король и ход за вами, то опять не важно, кто где стоит. Материал часто решает дело, поэтому оценка материала даже без учета его расположения может дать очень много.

Много, но не все. При игре королем против короля и ферзя вполне возможен пат, то есть ничейная ситуация. В русских шашках (поле 8×8) известно, что три дамки гарантированно ловят одну стандартной комбинацией, при условии что эта единственная дамка не стоит на главной диагонали. А значит, если противник, в свою очередь, может превратить свою единственную шашку в дамку, вставшую на главную диагональ, то ситуация из проигранной становится ничейной. Из этих примеров следует, что обе группы факторов – и материальные, и позиционные – могут решить исход игры, и обе группы должны

быть учтены самым тщательным образом. Но материальные факторы просчитать несколько легче, поэтому начнем с них.

У каждой из фигур есть собственная сила, не зависящая от расположения

К сожалению, в деле учета материала нет никакой строгой теории, и, может быть, такая теория даже невозможна. Ведь что нам нужно, по сути? Необходимо точно определить значимость каждой фигуры. Мы можем совершенно произвольно определить стоимость самой слабой фигуры – пешки или самой сильной – ферзя. Например, пусть пешка стоит 1 балл. А вот далее начинаются серьезные проблемы. Необходимо договориться, насколько сильнее пешки каждая из фигур. Причем безотносительно конкретной позиции. Это сложно во многих планах. Для любого шахматиста сила фигуры привязана к позиционным факторам: сдвоенные слоны сильнее двух несвязанных. Сила пешки увеличивается по мере ее продвижения к последней горизонтали. В дебюте ладьи мало что решают, но в эндшпиле, когда большая часть доски пустеет, их сила резко возрастает. Конь очень важен в миттельшпиле, но в эндшпиле ему сложно перескакивать с фланга на фланг. Слон скачет с большой легкостью, но у слона серьезные проблемы в миттельшпиле, где он на каждом шагу наталкивается на пешки и свои, и противника, кроме того, для слона доступна только половина шахматного поля – половина его цвета. Но тем не менее надо как-то от всего этого отвлечься и оценить собственную силу фигур, ведь интуитивно ясно, что есть какие-то внутренние характеристики.

Такая ситуация характерна для многофигурных игр, особенно если фигуры можно рубить. В русских шашках проблема проявляется слабее, видов фигур только два: шашка и дамка. В рэндзю проблемы материала вообще нет, в этой древней игре только один вид фигур – камень, и более того, после каждого хода количество камней игроков одинаково, а значит, материального фактора нет вообще. Но мы сейчас говорим о шахматах.

Стандартный подход к оценке материала опирается на составление экспертного мнения. Сделаем так: соберем группу опытных шахматистов и дадим задание оценить силу фигуры, например по десятибалльной шкале, приняв цену пешки в один балл. После чего составим оценку для каждой фигуры как среднее арифметическое оценок для данной фигуры по всем экспертам. Как это ни странно, но такой простой способ сработает, и довольно неплохо. Дело в том, что эксперты тоже не имеют никакой теории силы фигур, как и мы, а значит,

опираются на собственные интуитивные ощущения. В игре эксперт использует тоже свои представления о силе фигур, а следовательно, если эксперт добросовестно передаст свое понимание вопроса, он как бы передаст свой реальный опыт, а значит, даст программе возможность эффективно сыграть против себя. С учетом же того, что мы собрали сильных экспертов, становится ясно, что их личная сила, таким образом, передается и программе.

Но необходимо заметить, что вопрос, сформулированный так примитивно, – как оценить собственную силу фигур безотносительно позиции, – не дает эксперту возможности передать более значительную часть своего знания – знания о взаимной силе фигур.

Материальные факторы зависят от взаимного расположения, и это обстоятельство значительно усложняет анализ

Следующим шагом необходимо поинтересоваться у экспертной компании не только о том, как оценить некий фактор, но и какие вообще материальные факторы возможны. Имеются в виду факторы, определяемые положением фигур. Группа таких факторов выглядит намного сложнее. И здесь в полный рост встает проблема взаимопонимания. Цель разработчика – грамотно задать вопрос. Искусство вопроса предполагает, что задающий вопрос, может быть, не эксперт, но все же имеет хорошее представление о предмете. Поэтому хотите вы или нет, но если взялись писать программу, играющую в интеллектуальную игру, вам придется и самому научиться в нее играть на приличном уровне. Это даст возможность правильно сформулировать вопрос и выделить из ответов существенную информацию. Критерий хорошего экспертного ответа (в отношении определения фактора) таков: фактор должен быть легко алгоритмизируем. А это не всегда так. Вот пара примеров.

Хороший фактор. Сдвоенные слоны. Так называются слоны, находящиеся на соседних диагоналях. Это обстоятельство алгоритмически легко проверяется, поэтому фактор не только сильный с точки зрения шахматиста, но и удобный с точки зрения программиста.

Плохой фактор. Ладья имеет высокую степень свободы. Не вполне понятно, что имеется в виду. Этим может быть сказано, что ладья может своим ходом встать на значительное число полей, что, конечно, неплохо, но что, если все такие поля бьются фигурами противника. Кроме того, «высокая степень свободы» – это качественная оценка, а нам необходимы оценки количественные. Фактор можно превратить в хороший следующей переформулировкой: ладья полностью контро-

лирует открытую вертикаль. Что такое открытая вертикаль, понятно и легко проверяемо, а контроль означает, что ладья будет угрожать любой фигуре, вставшей на эту вертикаль, причем если вертикаль уже открыта, значит, на ней нет пешки. Любая же другая фигура сопоставима с ладьей, и, значит, ладейная угроза всегда будет существенна.

Если разработчику и экспертной группе удалось сформулировать хорошо алгоритмизируемые факторы материальной группы, то следующий шаг очевиден, необходимо опять провести численную оценку по той же шкале, по которой проводилась оценка абсолютной силы фигур.

Есть группа факторов, определяющих не силу фигуры, собственную или во взаимодействии с еще одной или двумя фигурами, а оценивающих качество позиции в целом. Это так называемые позиционные факторы

Здесь разработчика и экспертов ожидают наиболее существенные проблемы в связи с тем, что хорошо алгоритмизируемых позиционных факторов не бывает. Как, например, описать, что означает: давление на королевский фланг, контроль центра, низкая активность фигур, слабое фигурное взаимодействие на фланге и т. д.? Некоторые рекомендации и здесь возможны, но хорошее решение позиционной проблемы, обладающее математической строгостью, если и возможно, то пока даже в отношении такой разработанной игры, как шахматы, неизвестно.

В шахматах, однако, есть один хорошо алгоритмизируемый позиционный фактор – это давление на пункт. В отношении любого поля доски реально посчитать количество фигур, держащих этот пункт под ударом. Вот это количество ударов можно считать первичной оценкой фактора. Подчитав количество ударов на поля, принадлежащие центру, несложно оценить, кто из игроков имеет больше шансов выиграть борьбу за центр и, следовательно, кому из них плюсовать этот фактор. А оценку фактора, входящую в оценочную функцию, необходимо опять спрашивать у экспертной группы по уже известному сценарию.

Кстати, этот фактор – давление на пункт, или, иначе говоря, возможность захвата пункта, можно выделить в очень многих играх, а стало быть, его допустимо считать общеприменимым

Идея опоры на чисто арифметический подсчет ударов, конечно, весьма спорна, так как давление разными фигурами явно не равноценно. Например, не факт, что давление ферзем более сильно, чем

давление пешкой. В определенных условиях пешка в силу своей малоценности может оказаться более эффективной. Но, с другой стороны, учитывать в позиционном факторе еще и силу фигуры, осуществляющей давление, может слишком сильно усложнить оценочную функцию, увеличив вероятность ошибки. Здесь работает общее правило, утверждающее, что чем сложнее механизм, тем меньше вероятность его правильной работы.

Как считать оценочную функцию? Существует общий принцип, вытекающий из природы игровой стратегии, стремящейся минимизировать ущерб. Он называется принципом минимакса

Пока ясно, что есть факторы, описываемые двумя числами – ценой и количественным значением. Это как в овощном ларьке, у каждого овоща есть цена, и есть их наличное количество. Центр доски или королевский фланг, конечно, присутствуют на доске в единственном экземпляре, и пара сдвоенных слонов может быть только одна, но материальные факторы, в том числе и ферзь, могут быть в нескольких экземплярах (пешка может стать ферзем). Введем обозначения: m_k – это количество фактора, и v_k – это цена, или, еще говорят, вес фактора. Тогда общая оценка позиции может быть записана следующим выражением:

$$F = \sum_{k=1}^{N} m_k v_k = m_1 v_1 + m_2 v_2 + \ldots + m_N v_N .$$

Теперь попробуем разобраться, как оценочную функцию использовать для выбора хода. Для этого представим себе некую гипотетическую игру, в которой на каждый ход одного из игроков существуют ровно два ответа. Назовем игроков: ***Первый*** и ***Второй*** – и выберем продолжение для ***Первого*** игрока, при глубине анализа в один ход (один ход – это ход ***Первого*** и ответ ***Второго***). Анализ игры в этом случае будет опираться на такое дерево (рис. 2.4):

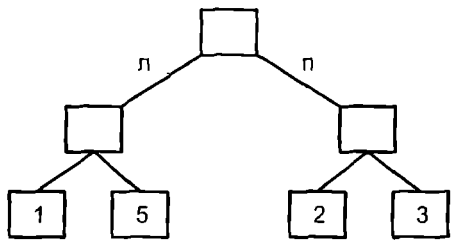

Рис. 2.4 ❖ Пример минимакса

Похожая картинка уже была, но здесь информации несколько больше. Конечные ситуации оценены, и есть вопрос, какой ход: левый или правый – должен выбрать ***Первый*** игрок. Возможно, кому-то покажется, что лучший ход – левый, так как именно там расположена ветка с наилучшей оценкой в пять баллов. Но нужно понимать, что после левого хода ***Первый*** игрок передаст право выбора хода ***Второму***, и этот ***Второй*** выберет ветку, оканчивающуюся единичной оценкой. Поэтому необходимо выбирать правый вариант, в котором ***Второй*** игрок отдаст ***Первому*** ветку с худшей оценкой 2. Заметим, что ***Первый*** игрок реально может выбирать только из худших оценок, и все, что в его силах, – это выбор наилучшего из худшего. Данный принцип называется принципом минимакса. Рассмотрим для иллюстрации более сложный пример. Дерево перебора оформим в виде таблицы, на нижнем уровне которой – значения оценочной функции на максимальной глубине дерева перебора, на промежуточных уровнях – значения оценки, выбранные противниками в свою очередь хода (оценки переходят снизу вверх):

<table>
<tr><td colspan="16">2</td></tr>
<tr><td colspan="8">1</td><td colspan="8">2</td></tr>
<tr><td colspan="4">4</td><td colspan="4">1</td><td colspan="4">4</td><td colspan="4">2</td></tr>
<tr><td colspan="2">1</td><td colspan="2">4</td><td colspan="2">1</td><td colspan="2">0</td><td colspan="2">4</td><td colspan="2">1</td><td colspan="2">1</td><td colspan="2">2</td></tr>
<tr><td>1</td><td>8</td><td>4</td><td>5</td><td>1</td><td>9</td><td>0</td><td>0</td><td>4</td><td>5</td><td>1</td><td>56</td><td>18</td><td>1</td><td>2</td><td>3</td></tr>
</table>

Как и любой эвристический механизм, минимакс имеет свои плюсы и минусы, не гарантируя наилучшего продолжения, он позволяет выбрать действительно разумный вариант игры

Принцип основан на предположении, что игроки имеют равную силу. Действительно, преуменьшать возможности оппонента нет смысла. Если окажется, что он играет сильнее, чем предполагалось, – это может иметь плохие последствия. Исходить из того соображения, что соперник играет лучше, – заведомо проигрышная психология. Оптимальная точка зрения заключается в том, что я играю так хорошо, как могу, и мой соперник играет не хуже. В терминах минимакса это означает следующее. Если выбор хода за мной, то я выберу ход с наилучшей оценкой для себя и наихудшей для противника, он, в свою очередь, сделает то же самое.

Из этого следует интересный вывод: если игра для каждого из игроков предоставляет равные возможности и если оценка позиции

содержит все возможные факторы и все они оценены верно, то игра для обоих игроков возможна только на ничью. А если один из игроков – компьютерная программа, не умеющая допускать ошибки по невнимательности, то у человека против такой программы нет шансов, так как в этом случае выигрывать будет тот, кто в состоянии проанализировать дерево перебора на большую глубину. Конечно же, счетные возможности компьютера неизмеримо выше, нежели человека. Но, к сожалению или к счастью, полный анализ достаточно сложной игры, даже такой, как русские шашки, представляет собой исключительно тяжелую теоретическую задачу, и похоже, такой анализ не был выполнен еще ни для одной игры с полной информацией. А значит, всегда есть возможность построить более качественную оценочную функцию.

Кроме того, всегда есть возможность принимать решение на основе опыта. Распознав ситуацию на доске, человек может найти в своей памяти прецедент, завершившийся положительным или отрицательным результатом после определенного хода, и этот прецедент даст информацию о ходе без анализа дерева перебора.

А теперь давайте посмотрим, что даст фактическая невозможность составить идеальную оценочную функцию. Она, идеальная функция, была бы не нужна, будь у нас возможность выстроить дерево перебора от начала игры до самого конца, в этом случае достаточна предельно простая оценка с одним фактором – игра выиграна или игра проиграна. Можно утверждать, что:

Чем глубже дерево перебора для игрока, тем меньше у него потребность в хорошей оценке. И наоборот, если построить идеальную оценочную функцию, то потребность в дереве перебора отпадет полностью. Очередной ход можно будет просто вычислять из знания текущей ситуации.

Но идеальная оценка невозможна, невозможно и полное или даже очень глубокое дерево перебора. Это означает, что вывод о качестве выбранного хода всегда может быть ошибочен, что создает возможность для так называемого комбинационного удара, выполнение которого выходит за рамки минимакса. В чем суть комбинации (тактического приема)? А суть в следующем: игрок допускает резкое ухудшение своей позиции в анализе дерева перебора на глубину в N ходов, но на большей глубине он получает значительно большую компенсацию. Например, можно отдать ферзя, если в результате противник получит мат, можно отдать легкую фигуру, если за этим последу-

ет взятие тяжелой фигуры противника. А иногда в шахматах отдают материал за позиционный выигрыш.

Заметим, конечно, что, увеличивая вычислительные ресурсы, мы даем возможность программе просчитывать варианты, которые выглядят как комбинации (отдача материала или позиции с последующим отыгрышем), но всегда есть несчитаемая глубина дерева перебора и всегда, с точки зрения минимакса, плохая оценка на максимальной глубине – это безусловно плохая оценка.

Оптимизация минимаксной процедуры. Альфа-бета-алгоритм

В этом параграфе мы обсудим два момента: как увеличить глубину дерева перебора и как без полного перебора обнаружить комбинационный удар. Начнем с метода, позволяющего более глубоко копнуть дерево перебора. Заметим, для начала, что более глубокое дерево без увеличения вычислительных ресурсов возможно только за счет отсечения некоторых его не очень значимых ветвей. Только это дает возможность другие, более важные ветки просмотреть глубже. А процедура отсечения ветки нуждается в критерии, позволяющем оценить ветвь игры до ее анализа.

Интуитивный критерий лежит на поверхности. Предположим, следующим ходом (напомню, мы в качестве базовой игры рассматриваем шахматы) игрок теряет ферзя. С точки зрения простой оценочной функции, это очень плохо, и такой ход разумно исключить из рассмотрения. Предположим далее, что следующим ходом противник теряет ферзя. Обычный здравый смысл говорит, что на этом варианте следует сосредоточиться. Но мы сделаем несколько парадоксальный вывод, что этого хода также следует избегать. Дело в том, что, выстраивая теорию, мы исходили из предположения, что силы игроков равны. Из чего следует, что ни один из игроков выиграть слишком много не может. И следовательно, как выигрыш ферзя, так и его проигрыш необходимо признать делом нереальным. Математически расчетная схема выглядит так: определим для оценочной функции пределы значений, для которых ни одному из игроков не гарантирована победа. Эти два уровня опять-таки можно определить экспертной оценкой. Назовем нижний уровень уровнем альфа, а верхний – уровнем бета. Далее все очень просто. Идя по дереву перебора, будем выполнять оценку промежуточных ситуаций (в чистом минимаксе интересны

только конечные позиции), и если эта оценка ниже уровня альфа или выше уровня бета, то такую игровую ветку отсекаем. На вопрос, как все же решиться взять ферзя противника, ответим немного позже.

Если вам показалось, что такой политикой игрок отказывается от возможности (но не любой, а лишь авантюрной, с точки зрения альфа-бета-алгоритма) выигрыша в пользу гарантированной ничьей, то вам показалось правильно. Цель алгоритма в том и состоит, чтобы обеспечить ничью при равных возможностях. Если же вы ставите своей целью победу, то необходимо несколько изменить подход.

Стратегическая борьба за победу и тактический удар вполне возможны, с точки зрения оценочной функции, но необходима более сложная схема расчетов

Выиграть партию можно двумя путями: постепенно, шаг за шагом наращивая преимущество или выполнив тактический удар, дающий сразу большое превосходство в оценке. Альфа-бета-алгоритм запрещает скачкообразное изменение оценочной функции, но оставляет возможность ее плавного роста. Действительно, в интервале между нижней и верхней границами лежат различные значения оценочной функции, и игрок в рамках минимакса будет стремиться приблизиться к верхней границе. Понятно, что чем ближе оценка текущей позиции к границе бета, тем больше будет появляться запретных ходов с точки зрения альфа-бета-алгоритма. Может даже возникнуть и совсем парадоксальная ситуация, в которой игрок, имея очень сильную позицию, не будет иметь ни одного хода, так как любой ход, улучшающий позицию, придется признать запретным, ибо любое улучшение оценки выйдет за границу бета. Единственный разумный выход из положения заключается в том, чтобы повышать границу бета с увеличением оценки текущей ситуации и понижать границу альфа с ухудшением игровой ситуации. Такой подход создаст возможность постепенного наращивания преимущества, которое может в некоторый момент стать решающим. И даже взятие ферзя в этом случае не будет противоречить алгоритму, если оценка бета достаточно высока.

Тактический удар, он же комбинация, также вполне алгоритмизируется в рамках изучаемой схемы, и достаточно легко. Границы альфа и бета – это границы возможной равновесной игры. А комбинация – это форма игры неравновесной, предполагающей резкие скачки оценки. Сказанное означает, что ход, выводящий оценку за границу, потенциально комбинационный. Обнаружив такой ход, программа может выполнить более глубокое построение дерева перебора именно

для такого хода, но ограничить его участком доски, для которого ход имеет существенное значение. Конечно, еще остается проблема, как оценить, для какого именно участка он существенен, но ответ на него зависит от системы оценки и природы игры.

Есть и принципиальная проблема. Заметим, что, определяя комбинационную возможность таким образом, мы фактически отказываемся от стратегии альфа-бета-алгоритма, так как все ходы, выходящие за границу, придется автоматически объявить потенциально комбинационными. Видимо, нужно уточнение критерия комбинации.

Для решения проблемы заметим, что факторы, входящие в оценочную функцию, не равнозначны. Между материальными факторами и всеми видами факторов позиционных есть существенное различие. Один и тот же материал можно иметь в разных позициях. Наличие пешек, легких и тяжелых фигур позиции не определяет. И наоборот, позиция заключает в себе игровые возможности, создаваемые особым расположением фигур. Именно расположение фигур дает возможность тактического удара. Это означает, что для уточнения, является данная ситуация комбинационной или нет, мы можем использовать какие-то позиционные моменты.

Для шахмат допустимо определить один позиционный фактор – давление на пункт или группу пунктов. Под давлением мы понимаем количество фигур, в том числе и пешек, способных при данном расположении поучаствовать в борьбе за пункт. Вполне правомерна такая гипотеза: если у игрока на данный пункт или часть игрового поля потенциал давления выше, то для него возможен тактический удар по захвату этого пространства и, как следствие, достижению материальных приобретений.

Заметим, что эта гипотеза имеет эвристическую природу. Против нее можно вполне разумно возразить. Например, если первый игрок атакует поле ладьей и ферзем, а второй игрок защищает это поле пешечной цепью, то чье давление сильнее, зависит не от количества угроз, а от структуры пешечной цепи. Существуют многочисленные и при этом более тонкие, нежели количество угроз, факторы, влияющие на успех атаки. Но чем хороша эвристика, она не дает точного ответа, она всего лишь из бесчисленного количества возможных продолжений выделяет наиболее вероятные, которые, конечно, нуждаются в проверке. И если программа обнаружит такой фактор – маячок, она может в этой части доски построить более глубокое дерево перебора, посредством которого получит более точную оценку возможности комбинационного удара.

Этапы игры

Еще одна интересная особенность, усложняющая анализ, – это изменение характера игры в зависимости от этапа. Партии очень многих игр, хотя и очень условно, можно разбить на три различных по природе этапа.

Есть начало игры. В стартовой позиции, в таких играх, как рэндзю, го, реверси, доска практически пустая, у обоих игроков нет ресурсов, достаточных для начала острой борьбы. В таких играх, как шашки, сеги, шахматы, все фигуры – на доске, но контакт с противником минимальный, а значит, серьезное нападение на построения оппонента невозможно. В играх, подобных войне башен (описание всех этих игр вы можете найти на lotos-khv.ru), уже в начале игры контакт между соперниками максимальный, но напряжение размазано по доске равномерно, ни у одного из игроков нет перевеса в какой-либо части поля, а значит, решающая атака первые 2–3 хода невозможна. Это означает, что в начале игры противники реально могут ставить перед собой только стратегические задачи по накапливанию позиционного преимущества. В шахматах этот этап называется дебютом.

Есть активная фаза игры. Здесь уже возможно все, и стратегическая борьба, ставящая целью улучшение позиции, и тактические удары, направленные на получение существенного позиционного или даже материального превосходства. В этой фазе имеет место полный контакт во всех частях игрового поля, и в идеале весь материал опытного игрока участвует в борьбе. Эта фаза игры отличается быстрой переменой ситуации, необходимостью учитывать большое количество факторов, собственная сила фигур становится фактором очень шатким, удачно стоящая пешка может быть сильнее ферзя и т. д. Этот этап в шахматах называется миттельшпилем.

В любой игре есть завершающая фаза. Она не всегда сводится к добиванию соперника. Достаточно часто и здесь идет равная борьба. Но, например, в шахматах или шашках концовка характеризуется малым количеством материала, а значит, собственная сила фигур приобретает большое значение. На этом этапе ферзь – это уже ферзь во всей своей красе и мощи. Большую силу приобретают и факторы расположения фигур. Здесь уже стратегия значит больше тактики, и хотя комбинационные удары и тонкая тактическая игра по-прежнему возможны, накал борьбы все же спадает. В рэндзю или реверси большие участки игрового поля выходят из игры, и борьба сосредотачивается в отдельных очагах активности. В шахматах этот этап называется эндшпилем.

Понятно и без длинных рассуждений, что описанные различия в этапах игры настолько принципиальны, что их необходимо учитывать в любой формализованной схеме игры. В том числе и в нашей технологии, построенной на оценочной функции. В первой прикидке ясно, что оценочную функцию нельзя построить на все времена, и для каждого из указанных этапов она должна быть немного иная. Попробуем дать некоторые рекомендации.

Дебют

В некотором смысле самая сложная фаза. В шахматах проблема дебюта решается машинными ресурсами. Дело в том, что шахматы – это игра с хорошо развитой дебютной теорией. Настолько хорошей, что знание теории обеспечивает в дебюте решающее преимущество, поэтому вполне достаточно сделать полное формализованное описание этой теории. Задача чисто техническая и неинтересная. Давайте подумаем, а что бы мы стали делать, если бы теории не было. Тем более что в большинстве игр ситуация обстоит именно так – нет никакой теории.

Прежде всего дебют отличается специфической игровой целью. Заметим сразу, что в шахматах отыгрыш материала в дебюте, даже одной пешки, – это очень серьезно. Отдача материала в миттельшпиле может означать комбинационный удар, но в дебюте таких возможностей, как правило, нет, поэтому потеря любой фигуры в начале игры, скорее всего, восполнена не будет, поэтому для дебютной оценочной функции можно не выделять фигуры как отдельные факторы. Целесообразно выделить два материальных фактора: пешечный, уменьшение которого разумно считать крайне опасным (но не фатальным, мы помним о возможности гамбитов), и фигурный, уменьшение которого можно признать сопоставимым с поражением в партии.

Вернемся к определению специфической игровой задачи. Шахматисты, говоря о дебютной цели, используют термин «развитие». Это означает, что в начале игры необходимо обеспечить максимальную подвижность фигур, отсутствие уязвимых пунктов в позиции и захват стратегически важных территорий. Как учесть эти факторы?

Подвижность фигур, очевидно, не распространяется на пешки. Их атакующий потенциал не очень высок даже в сравнении с легкими фигурами, поэтому пешки, скорее, играют роль оборонных пунктов, поддерживающих наступательные действия. И даже если пешки участвуют в атаке, и тогда они не вполне самостоятельны. Кроме того, их подвижность резко ограничена правилами хода. Точно так же нель-

зя считать плюсом позиции высокую подвижность короля. В дебюте, при большой насыщенности доски фигурами, для короля слишком много угроз, для того чтобы он стал полноценной играющей фигурой. Есть проблемы и с ладьями. Во-первых, ладья – очень ценная фигура, но даже не это главное, еще более ценный ферзь – активный участник дебюта. Проблема ладьи в том, что она не умеет преодолевать пешечную цепь. Поэтому в начале игры ладья объективно сильно связана в пространстве пешками и противника, и своими.

И только три типа фигур способны активно перемещаться в первой фазе игры: это слоны, кони и ферзь. Подвижность этих фигур и надо стремиться нарастить. Сам же фактор подвижности можно измерять двумя способами. Например, учитывать все поля, на которые фигура способна встать в принципе, но, вероятно, точнее было бы учитывать не битые поля, а лишь те, на которые фигура может встать без угрозы погибнуть.

Поговорим об уязвимости позиции. Речь, конечно же, идет обо всей позиции, контролируемой игроком. Но учитывать атаку на подвижные легкие фигуры или ферзя в плане контроля пространства большого смысла нет в силу их подвижности. Или этот учет будет довольно сложен. Поэтому разумно ограничиться анализом слабостей пешечной цепи и так называемых стратегически важных пунктов, дающих контроль части доски, на которой находятся существенные собственные силы или силы противника. Количественную же оценку выполнить достаточно просто. Можно просчитать локальное (относящееся только к данной части доски) дерево перебора и выяснить, кто сможет овладеть этим пунктом с меньшими потерями, при условии что борьба будет вестись всеми доступными ресурсами только за него.

Этим замечанием мы завершим обзор дебютных идей и двинемся дальше, перескочив миттельшпиль. Дело в том, что все, что в этой главе уже было сказано о построении оценочной функции, дерева перебора, поиска комбинационного удара, – это все приложимо именно к миттельшпилю. Середина игры – это, так сказать, наиболее характерная часть партии. В эндшпиле же опять появляются особенности, которые необходимо учесть.

Эндшпиль

В завершающей части игры меняется вес различных факторов. Во-первых, король (особенно в отсутствие ферзей) становится реальной боевой силой, и достаточно грозной. Во-вторых, резко возрастает роль

пешек. Если глубоко продвинутая пешка в миттельшпиле – серьезный стратегический фактор, разрушающий позицию соперника и создающий угрозу на будущее, то в эндшпиле пешка на шестой и даже пятой горизонтали создает возможность быстрого тактического удара, завершающегося приобретением дополнительной фигуры. В-третьих, серьезно меняется соотношение между конем и слоном. В миттельшпиле конь как наиболее маневренная фигура, способная перескакивать через пешечные построения, сильнее слона, в эндшпиле же, при игре на два фланга, способность слона одним ходом переходить с фланга на фланг резко повышает ценность этой фигуры. И в-четвертых, ладьи, наконец, выходят на оперативный простор, превращаясь, образно говоря, из тяжелой артиллерии миттельшпиля в танковые войска эндшпиля. Ферзь, особенно при открытом короле противника, приобретает особую мощь.

Таким образом, оценочную функцию для эндшпиля необходимо строить заново, для всех материальных факторов. Не менее серьезно изменяются и факторы позиционные. Центр зачастую утрачивает свою доминирующую роль, нет уже ни королевского, ни ферзевого фланга, как в дебюте или миттельшпиле. Резко уменьшаются возможности для комбинационного удара. Нет в эндшпиле и общей для всех партий задачи развития, как в дебюте.

В силу упрощения ситуации на доске изменяется задача планирования. А именно возникает возможность узконаправленного плана, например можно поставить цель матовой атаки на короля или организации прохода пешки до последней горизонтали. Узкие задачи возможны и в миттельшпиле, но в эндшпиле больше возможности сосредоточиться на специальной проблеме.

Примечание. Напомним, что наше изложение использует шахматы только в качестве хорошего примера. А все сказанное может быть проецировано и на другие игры. Например, эндшпиль. Практически в любой игре на досках к концу партии значительные пространства выходят из игры, как, например, в рэндзю, где большие пространства могут быть сплошь заставлены камнями. Тогда для разных частей доски рэндзю, где еще идет борьба, возможны свои независимые планы.

Тогда есть смысл в оценочную функцию ввести фактор, численно описывающий именно приближение к этой цели. Такой фактор многое меняет. Например, потеря слона, дающая возможность пешки уйти вперед, может считаться компенсирующим плюсом.

Кроме того, именно в эндшпиле существуют так называемые ничейные ситуации. Например, король и ферзь не выигрывают против короля и ладьи (и на доске больше ничего), несмотря на то что относительная сила ферзя значительно выше ладейной. Нельзя поставить мат одним слоном или одним конем (на доске больше нет ни фигур, ни пешек). Есть ситуации, когда король не имеет возможности провести свою пешку. Таким образом, есть ситуации, когда доминирующее материальное превосходство уже ничего не дает, и вполне допустимо идти на потери, компенсация которых – гарантированная ничья. Есть ситуации, которые не столь однозначны, например один из противников имеет преимущество, позиционное или материальное, но даже глубокий анализ не выявляет возможности реализовать это преимущество. Эндшпиль – это та часть игры, в которой достижение преимущества может оказаться недостаточным для завершения партии победой. Этого обстоятельства нет ни в дебюте, ни в миттельшпиле. Мы показали различие между тремя этапами игры на примере шахмат, но нечто подобное можно выделить в любой игре, а значит, в любой игре надо ориентироваться на необходимость построения трех различных оценочных функций.

Оценка, основанная на приоритетах факторов

Способ построения оценочной функции, рассмотренный выше, не единственно возможный. Сейчас разберем несколько иную технологию, для которой справедливо все вышесказанное, но выглядит способ счета оценки существенно иначе. Анализ опять проведем на шахматном материале.

Заметим, что есть в шахматах события, настолько сильно влияющие на ход партии, что их можно признать доминирующими. Например, выигрыш или проигрыш ферзя настолько сложно скомпенсировать, что это событие можно признать почти равнозначным победе (поражению). Таких ситуаций достаточно много. Например, проигрыш центра в дебюте – практически некомпенсируемая потеря, лишняя пешка в эндшпиле при прочих равных также становится доминирующим фактором. А выигрыш фигуры в эндшпиле способен с лихвой уравновесить пешечную недостаточность. В общем, идея такова: есть факторы, влияние которых на ход игры настолько велико, что их превосходство позволяет не считать остальные.

Приняв такую доктрину построения оценки, мы создаем не одну оценочную функцию, а целую группу, внутри которой оценки упорядочены по значимости. При таком подходе счет оценки несколько упрощается. Доминирующие факторы высшего уровня обычно будут представлять из себя нечто, достаточно грубое и легко считаемое. Например, в дебюте самым значимым фактором можно признать перевес в материале (даже в пешках), что считается очень просто.

Таким образом, посчитав наиболее приоритетные оценки для существующих вариантов, программа может переходить к счету оценок следующего по приоритету уровня только в том случае, если оценки данного уровня приоритета не дадут достаточной информации для выбора хода.

Все остальное, что было сказано выше, и изменение ценности факторов с переходом в миттельшпиль или эндшпиль, и необходимость специальной техники анализа для поиска тактического удара, – все это остается в силе. Заметим только, что такая технология построения оценки более приближена к человеческому способу принятия решения и, следовательно, упрощает работу с экспертной группой.

Интересная гипотеза. Интегральный признак

Идея оценочной функции основана на возможности перечисления факторов, влияющих на ход игры, и оценки их значимости. Констатация того обстоятельства, что факторы имеют разное значение, наводит на мысль, что, может быть, существует один-единственный фактор, определяющий ход игры исчерпывающе.

В играх позиционных, таких как рэндзю, или го-бан (игра, очень сильно похожая на рэндзю), идея существования такого фактора почти очевидна. Действительно, в этих играх камни противника не забираются, нет правила рубки, что почти исключает комбинационные возможности. На доске количество камней одного из игроков всегда отличается от другого только на один, а значит, материальные факторы отсутствуют. Отсюда можно предположить, что в таких играх решающим фактором становится размер контролируемого пространства. И действительно, большее пространство создает возможности для развития позиции, что приводит к росту вероятности поставить пять камней в ряд. Кстати, и сама задача постановки в один ряд такого

большого количества камней предполагает наличие у игрока значительного свободного пространства.

Конечно, здесь возникает вопрос: что считать контролируемым пространством? Если правила рэндзю не предполагают рубки, а такой возможности, как уже было сказано, нет, то все поля, доступные для одного из игроков, очевидно, доступны и для другого. Пространственный фактор сомнителен и в другой позиционной игре – го. Сама цель игры го формулируется как захват как можно большего пространства. Объявлять же цель игры фактором игры несколько бессмысленно.

Видимо, для того чтобы можно было пространство объявить единым игровым фактором, к нему надо что-то добавить. В отношении рэндзю такая доформулировка выглядит достаточно просто. Будем считать полезным пространством такое, на котором игрок может создать угрозу. Напомним, что в этой игре угрозой считаются открытая тройка (ряд камней, способный следующим ходом превратиться в четверку, пусть даже прикрытую) и любая четверка (ряд камней, способный следующим ходом превратиться в пятерку и таким образом завершить партию победой). Рассмотрим следующий пример (рис. 2.5):

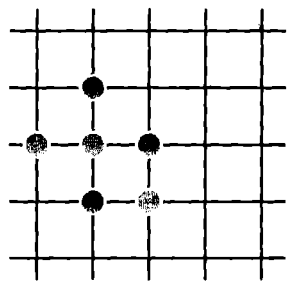

Рис. 2.5 ❖ Позиция рэндзю

В этой позиции синий игрок имеет возможность создать две различные угрозы, и каждая из них может быть реализована двумя различными способами, а зеленый – только одну, также двумя различными способами. Следовательно, можно утверждать, что синие контролируют большее пространство. Можно даже пойти немного дальше и посчитать, сколько угроз сможет создать каждый из игроков в течение нескольких ходов, если играть будет только он (безответно со стороны противника). Может быть, подсчитанное количество и не

определяет ход игры исчерпывающе, но, согласитесь, идея интересная и достойна тщательного анализа.

Конечно, в играх, предполагающих комбинационное развитие, найти такой интегральный фактор сложнее. Но заметим, однако, что комбинация не возникает на пустом месте, ее предваряет успешная позиционная игра, создающая какие-то скрытые возможности.

Давайте разберемся детальнее с термином «комбинация». Рассмотрим абстрактную игру на стандартной шахматной доске. Два игрока, будем называть их ***Первый*** и ***Второй***, ведут борьбу фигурами одинаковой ценности, и пусть в некотором игровом эпизоде они оказывают давление на два пространственных пункта: А и В. Под давлением на пункт будем называть количество ударов (возможностей срубить то, что стоит на этом пункте). Предположим, что давление со стороны каждого игрока на каждый пункт одинаково. Это означает, что ни один из них не сможет выиграть борьбу, если начнет ее первый. Но пусть у ***Первого*** фигуры, бьющие пункт А, и фигуры, бьющие пункт В, – это разные фигуры, а у ***Второго*** некоторые из фигур участвуют в давлении на оба пункта. Тогда со стороны ***Первого*** возможен комбинационный удар. А именно, атаковав один из этих пунктов, он вынуждает ***Второго*** отвлечь силы от защиты другого пункта, на который ***Первый*** и направит решающие усилия.

Этот пример наводит на мысль, что в играх комбинационного плана также возможен чистый счет одного признака. Например, можно посчитать для каждого важного пункта доски (или вообще для каждого пункта), кто выиграет это игровое поле, если вести борьбу только за него. Тогда стратегический перевес остается за тем игроком, который в состоянии потенциально обеспечить себе перевес в пространстве. Пространственное превосходство влечет за собой потенциальную возможность превосходства и материального.

Еще интересна эта идея тем, что анализ общей ситуации сводится к анализу локальных стычек на каждом поле доски, без учета того, как будет вестись борьба на других полях, что дает возможность отвлечься от практически необозримой системы взаимосвязей.

Есть в идее интегрального фактора, основанного на просчете локального давления, два существенных недостатка, делающих идею сырой. Во-первых, исходное предположение об одинаковой силе фигур – очень грубое предположение. Есть много игр с самыми разными фигурами. И согласитесь, если в шахматной позиции ***Первый*** игрок атакует некое поле ферзем и ладьей, а ***Второй*** защищает это же поле только одной пешкой, то локальная позиция ***Второго*** сильнее, так

как атака ***Первого*** будет означать для него слишком большие потери. Эту проблему можно решить, учитывая понятие силы фигуры. И мы уже достаточно обсуждали проблемы, связанные с относительной силой, – это и различные возможности фигур в разных стадиях игры, и сложности, связанные с учетом их взаимодействия, и ряд других проблем, делающих задачу оценки фигурной силы очень непростой.

Второе слабое место заключается в понимании важности игрового пункта. Ясно, что поля игровой доски неравнозначны. Насколько они неравнозначны? Как захват того или иного поля повлияет на продолжение борьбы? Более того, захват некоторых полей вполне может изменить значимость других полей и значимость имеющихся потенциальных угроз.

В общем и целом идея конструирования единого признака лично мне представляется весьма интересной. В моей школе программистов одна из первых игр, сделанных моими учениками, – «война башен» – была реализована на базе именно этой идеи. Начиналась работа с создания довольно объемной оценочной функции, затем, выявляя взаимосвязи факторов, мы отбрасывали их один за другим, в результате получив однофакторную функцию, играющую против человека весьма прилично. После, к сожалению, в стремлении улучшить качество игры оценка опять несколько усложнилась, но это не дало усиления игры. И программу, и описание игры желающие могут найти на www.lotos-khv.ru. Эта практика показала, что реально однофакторная функция потребует глубокого понимания игры, и без какой-то серьезной теории, расшивающей два упомянутых проблемных пункта, не обойтись.

В заключение

Наш анализ построения игровых программ посредством технологии оценочной функции и дерева перебора в основном велся на шахматном материале как наиболее знакомом для большинства потенциальных читателей этой книги. Но общие идеи, очевидно, будут работать для любой игры, и чем игра проще, тем, видимо, технология сработает эффективнее. Есть даже игры, чьи правила и игровое пространство позволяют построить исчерпывающее дерево перебора, а значит, обойтись минимальной оценочной функцией, может быть, даже только с двумя значениями: выиграл, проиграл. Но по-настоящему интересные игры, конечно, такого не позволяют. Как максимум, можно полностью просчитать завершающую фазу игры. Например, это

можно сделать в реверси, игре, изобретенной в Германии, и мельнице, пришедшей из древней Средней Азии. *Кстати*, за счет такой возможности именно реверси стала популярна для разработчиков игровых программ. Имея возможность просчитать последние 10–15 ходов, вполне *допустимо проявить некоторую* небрежность в середине игры, с учетом, что противник – человек – скорее всего, также будет несколько небрежен. Для сложных игр, повторимся, возможна гипотеза *идеальной* оценочной функции, позволяющей делать вывод о ходе без просмотра дерева перебора, *но это не более чем* гипотеза, пока что ни для одной игры подобная эффективно работающая функция построена не была.

Еще заметим, что технология «оценочная функция плюс дерево перебора» совершенно не человеческая. А это означает, что возможны и другие технологии. Там, где есть два варианта, там, скорее всего, будет и третий. А где есть третий, там, не исключено, есть и четвертый. Но моей целью не было *построение неизвестного*, поэтому на этом я главу и завершу.

Глава 3

Интеллект искусственный и обучаемый

Было бы очень интересно понять, насколько наш разум универсален. Может ли существовать интеллект, отличающийся от нашего не степенью развития, а организацией мыслительного процесса, методологией мышления? В уровне развития различия мы можем найти и внутри человеческой расы. Например, есть существенное различие между доктором математики и дикарем из Центральной Африки. Но если взять младенца – от родителей дикарей, поместить его в цивилизованную среду, дать хорошее воспитание, образование, то, видимо, результат будет таким же, как и для ребенка, уже рожденного в цивилизованной стране. Собственно, дикари от жителей Европы, Китая или любой другой развитой страны отличаются только объемом и качеством интеллектуального опыта. А это отличие количественное, а не качественное.

И здесь возникает интересный вопрос. А что такое качественное отличие? Как всегда, нужен тест, позволяющий определить, что мы видим перед собой. Например, обязан ли иной интеллект в своем развитии прийти к созданию теории относительности и дифференциального исчисления? Обязательно ли первые астрономические знания возникают из потребности ориентировки в пространстве, а геометрия – из земледелия? Будет ли иной разум доказывать те же теоремы, что и люди? Абсолютна ли та система наук, которую мы сейчас имеем, или есть альтернатива? Насколько абсолютен тот метод исследования, который мы называем научным? Вопросов такого рода

можно задавать бесконечно много. И чтобы на них ответить, нужен хотя бы один пример.

Вообще, на планете Земля существуют интеллектуальные системы, качественно отличные от человека. Это животные. Не знаю, кто как, а лично я верю, что коты, собаки и все прочие четвероногие и даже безногие обладают и душой (что бы это не означало), и разумом, но, к сожалению, эти разумные системы слишком примитивны, и их использовать в качестве примера нельзя.

Но если мы не можем назвать возможных отличий между нашим разумом и гипотетическим интеллектом, то одно общее сходство существует. Очевидно, любой разум должен иметь свойство обучаемости. Не берусь утверждать, что любого человека можно обучить любой области знания в совершенстве, но очевидно, что любой человек в любой области может воспринять какой-то объем знаний и умений. Если, например, отдельно взятого индивидуума не получится обучить счету интегралов, то, во всяком случае, научить арифметике можно. Если человек окажется не в состоянии стать полиглотом, то хотя бы научить его читать со словарем с одного родственного ему языка можно и т. д.

То есть человек не просто обучаем –
он универсально обучаем

И это свойство если и не является определяющим для интеллекта, то, во всяком случае, его можно признать обязательным. Попробуем проанализировать хотя бы кратко, в чем заключается наша способность к обучению.

Проблемы построения обучаемых систем

Систему искусственного интеллекта можно создать сразу, задав оптимальную стратегию поведения, научив ее вычислять правильные действия. Как это делать, мы рассмотрели в предыдущей главе. Сильная сторона такого подхода заключается в том, что не нужен очень умный компьютер. Все проблемы моделирования поведения программист берет на себя. Слабая сторона заключается в том же. Созданная система имеет ограничение по интеллекту, она – всего лишь модель, и любое улучшение требует модификации программы разработчиком. Более интересна возможность создания изначально слабой модели, но способной к саморазвитию. То, что это возможно, мы уже знаем на примере собственного разума, а сейчас рассмотрим некоторые вопросы, требующие разрешения на этом пути.

Существует универсальный метод познания, не гарантирующий серьезного результата, но применимый с чистого листа, – это метод проб и ошибок

Быть обученным – это значит с уверенностью знать, что означает то, что ты видишь, и что последует за тем или иным действием. Например, мы знаем, что если чиркнуть спичкой по шероховатой поверхности, то спичка, скорее всего, загорится, если сложить два и два, то обязательно будет четыре и т. д. Обученность означает владение знанием. Вопрос в том, как это знание можно получить. Вернемся к спичке. Мы хотим узнать, от какого действия она может загореться. Этот вопрос можно задать знающему человеку, прочитать где-нибудь (авторитетный источник), но сначала, для того чтобы задать этот вопрос, мы должны знать, что спичка вообще может загореться. То есть для того, чтобы получить новое знание, необходимо уже что-то знать, чтобы задать окружающей информационной среде грамотный, корректно сформулированный вопрос. А что, если нет никакой первичной информации? Или, что бывает чаще, ситуация настолько нестандартна, что аналогий, позволяющих привязаться к какому-то опыту, нет. В этом случае выход только один: **если ты не знаешь, что делать, надо делать что-нибудь**.

Действие в неизвестной среде вызовет ответную реакцию (будем надеяться, что эта реакция не приведет к летальному исходу) и даст информацию, в худшем случае негативного характера. Но отрицательный ответ, как говорится, тоже ответ, и таким образом исследователь начнет поиск решения поставленной задачи, пробуя различные варианты действия и отсеивая ошибки. Такой метод называется методом проб и ошибок. Он, конечно же, неэффективен, но в отсутствие какой-либо первичной информации это единственная возможность накопления знания.

Такой способ доступен любым интеллектуальным системам. Зоологи говорят, что таким образом стая крыс собирает знание о съедобной пище. В стае всегда есть несколько камикадзе, готовых рискнуть своей жизнью, для того чтобы проверить качество неизвестной пищи, а результат такой проверки, удачный или неудачный, запоминает вся стая. Собственно, этот способ можно назвать базовым и для человека. Все индуктивные, дедуктивные методы, наша способность к абстракции, обобщению и т. д. начинаются отсюда, с метода проб и ошибок.

Метод проб и ошибок выращивает хлеб разума – опыт

Метод проб и ошибок представляет собой исходный источник опыта для вида, но не для каждого отдельно взятого представителя популяции разумных существ. Уже ясно, что лично набивать шишки, исследуя мир, – это слишком накладно. Поэтому природа дала всем умеющим думать способность к обмену информацией как компенсацию за слабый исследовательский старт. Опыт, приобретенный одним, становится общим достоянием. Однако, для того чтобы опытом можно было воспользоваться, необходимы довольно сложные инструменты. Требуется средство для коммуникации – так называемый язык. И мы можем заметить, что человеческий язык не просто сложнее алгоритмического, он качественно иное образование. Если машинные языки, так, как мы их понимаем, являются только лишь передатчиками команд, то человеческий содержит мощный понятийный аппарат, с помощью которого мы даем наименования всем внешним сущностям и описываем связи между ними.

Еще одна вещь, необходимая для получения и использования опыта, – это память. Причем не просто способность запоминания цепочек битов или байтов, а специальным образом организованная память, предназначенная для хранения сложноструктурированной информации. В современной теории искусственного интеллекта такая память называется базой знаний. Постарайтесь не перепутать с базой данных. Немного позже мы рассмотрим понятие знания и базы знания детальнее и выясним различие с просто данными.

Накопление опыта дает возможность менять стратегию поведения. Опыт делает деятельность, в том числе и исследовательскую, более успешной. Если научиться накапливать опыт и использовать его, то «поумнение» происходит взрывом. Накопление знания увеличивает наши возможности в дальнейшем накоплении, и этот процесс непрерывно ускоряется. По сути, именно это явление мы и наблюдаем на примере развития нашей, человеческой цивилизации. Но в интеллектуальных системах, способных использовать опыт, мы наблюдаем еще одну, даже более интересную вещь. Приобретя некоторый опыт, мы на самом деле знаем больше, чем в этом опыте непосредственно заложено.

Получив даже небольшой опыт, мы знаем больше в силу нашей способности вывода нового знания чисто умозрительными методами

Для начала простой пример. Перепрыгнув через данную конкретную лужу, мы узнаем не только то, что через эту лужу можно пере-

прыгнуть. Мы узнаем, что можно перепрыгнуть через некую обобщенную лужу, а значит, получаем знание о способе перепрыгивания через лужи вообще. Более того, мы узнаем, что можем перепрыгивать, и лужа как препятствие абстрагируется в препятствие вообще. Пример, конечно, упрощенный, но для иллюстрации общей способности разума вполне годится. Разум способен интерполировать полученную информацию на похожие ситуации.

Интерполяция – очень интересный метод, дающий возможность формулировать правдоподобные эвристические гипотезы в ситуациях, для которых есть аналогии. Интересно здесь то, что процедурой обобщения описанную возможность объяснить нельзя. Обобщение требует опыта во множественном числе, обобщенный опыт предполагает, что было отработано значительное количество похожих ситуаций. Здесь же разум, анализируя единичную ситуацию, ищет некое ключевое сходство, достаточное для применения опыта. Интерполяция представляет собой способность предположения на основе опыта, который накладывается на новую реальность, без множественного закрепления удачными исходами.

Игра как проблема обучения

Конечно, системы искусственного интеллекта, построенные на умении программы обучаться, будут очень и очень интересны тогда, когда они смогут перенастраиваться с задачи на задачу. Возможность перенастройки требует решения очень сложных проблем, сложных даже для задач искусственного интеллекта. Например, задача игры и задача управления сложной технической системой требуют совершенно различных систем базовых понятий и методов принятия решений, и этот переход должен происходить в рамках одной и той же программной технологии. Проиллюстрируем проблему на яблоках и картошке.

Предположим, некий изобретатель построил две машины: одна собирает фрукты с деревьев, другая копает овощи из земли. Ясно, что обе эти машины конструктивно будут отличаться. Затем эти машины начали развиваться (людьми, конечно). Для первой появились модификации, собирающие: бананы, яблоки, мандарины и т. д. Все они будут иметь конструктивные особенности при общей идейной базе. Для второй появились модификации, копающие: картошку, свеклу, морковь и т. д. Ясно, что в этих машинах тоже появятся технические особенности при наличии общей конструкторской идеи. Потом все

машины первого типа объединили в машину *A*, способную перенастраиваться на разные фрукты, растущие на дереве, а все машины второго типа – в машину *B*, способную перенастраиваться на разные корнеплоды. Ясно, что машины *A* и *B*, переходя от задачи к задаче, меняют технические особенности в рамках общей конструкции. А вот если мы пожелаем объединить эти две машины в некую машину *C*, способную собирать яблоки и копать картофель, то либо мы чисто механически прикручиваем *A* к *B*, что, конечно, плохой вариант, либо создаем машину, способную радикально изменять свою конструкцию при переходе от яблок к картофелю.

Смена, по ходу работы, понятийного аппарата, конструктивных решений – это одна задача. Есть еще проблема рисков. Например, для индивида нет возможности накапливать опыт прыжков через пропасть, прыгая через разные пропасти с запоминанием ситуаций, в которых это не удалось, нельзя так моделировать обучение машины и в управлении критически опасными техническими системами, например атомной станцией. В общем случае необходимо понимать, что в реальных, жизненных задачах накопление опыта зачастую идет с риском приобретения необратимых последствий. Но переход от известной задачи к совершенно новой происходит в условиях нулевого опыта, а, как мы уже говорили, в такой ситуации первичное накопление опыта возможно только методом проб и ошибок.

Здесь систему искусственного интеллекта может спасти только умение интерполировать имеющийся минимальный опыт на новую ситуацию, для того чтобы как-то минимизировать риски. Минимальный опыт мы можем себе позволить. Например, человеческий младенец не входит в жизнь с чистой памятью. Можно предположить, что инстинкты и простейшие рефлексы – это минимальная память, накопленная всеми живыми существами (даже не только людьми), довольно значительный опыт ребенку передают родители, и это так у всех высокоразвитых животных. Усвоив эти важные моменты, перейдем к игре.

Интеллектуальные игры хороши тем, что в их рамках можно позволить себе простейшую технологию интеллектуального развития – метод проб и ошибок. Проигранная партия дает серьезную информацию и не несет в себе угрозы существования интеллектуальной системы.

Как научить машину учиться игре

Начнем с описания простой схемы обучения. Пусть есть некая игра. Известно эвристическое правило, позволяющее определять очеред-

ной ход. Например, это некая оценочная функция. В оценочную функцию входит ряд коэффициентов, значения которых можно изменять. Мы предполагаем, что изменение коэффициентов влияет на качество игры, но не знаем, как. Возьмем две программы. Одну назовем ***Альфа***, другую ***Бета***. ***Альфа*** имеет возможность произвольным образом изменять значения коэффициентов, ***Бета*** использует одну и ту же функцию, переданную ей ***Альфой***. ***Альфа*** будет играть с ***Бетой***, изменяя оценочную функцию, в случае успеха закрепляя изменения и в случае неудачи их отменяя.

Альфа в начале эксперимента передает ***Бете*** исходную функцию (любую) и изменяет коэффициенты в своей версии. Затем обе программы играют серию игр. Если ***Альфа*** статистически неуспешна (чаще проигрывает), то изменения, внесенные ***Альфой***, считаются неудачными. Если ***Альфа*** успешна, то ее версия оценочной функции передается ***Бете*** в качестве нового стандарта. Это общая схема, в которой необходимо очень многие вещи детализировать. Чем мы сейчас и займемся. Прежде всего определим некоторые полезные ограничения, которые сделают задачу решаемой, на достаточно интересном, но не слишком сложном уровне. Кто забыл, почитайте все, что было сказано об оценочной функции в предыдущей главе. Дальнейшие исследования идей обучаемости машины будем выполнять на этой конструкции:

$$F = \sum_{k=1}^{N} m_k v_k = m_1 v_1 + m_2 v_2 + \ldots + m_N v_N.$$

Альфа и ***Бета*** на начало учебного процесса имеют одинаковые версии оценки, затем ***Альфа*** должна по некоей стратегии изменять коэффициенты игровых факторов, это как минимум. Более продвинутая ***Альфа*** должна иметь набор факторов и стратегию более радикального изменения оценки путем исключения факторов, не оправдавших надежды и включения из библиотеки других. Эти две стратегии (изменения весов факторов и изменения вида функции) мы должны обсудить.

Проблема подбора весов решалась бы легко, если бы между весом фактора и качеством игры ***Альфы*** была прямая зависимость. Если фактор улучшает качество игры, тогда чем вес фактора больше, тем оценка лучше. Такие простые факторы даже для шахмат построить несложно. Наверное, можно показать, что такая система факторов возможна для любой игры. Но возможность наращивать вес фактора ограничивается ресурсами игрока. Например, нельзя бесконечно

наращивать давление на центр шахматной позиции, как максимум возможное давление ограничено количеством фигур, а в реальности и эти ограниченные ресурсы необходимо распределять между различными задачами, каждая из которых может быть представлена соответствующим позиционным фактором. Нельзя все ресурсы бросить в атаку на ферзевый фланг и пропустить решающий удар по своему королю. Нельзя сосредоточить все силы на обороне короля и позволить противнику откусывать пешки одну за другой.

Говоря языком математики, различные факторы взаимосвязаны, и игнорировать эту связь, даже в грубой модели, нельзя. В необучаемой программе проблемы взаимосвязи, взаимозависимости факторов мы возложили на экспертов, после работы которых машине передается уже качественная оценка. Сейчас мы должны научить машину обойтись без экспертной помощи.

Доминирующие факторы

Таким образом, главная проблема ***Альфы*** заключается во взаимосвязанности факторов. Изменение только одного из них не дает никакой информации о взаимозависимости с другими. В результате машина вынуждена делать выбор наугад, то есть в чистом виде использовать метод проб и ошибок. Но тогда обучение становится делом, чисто вероятностным. Предположим, ***Альфа*** увеличила вес фактора *B*, и изменение оказалось успешным. Означает ли это желательность его дополнительного увеличения? Конечно нет. Означает ли это, что, наоборот, его уменьшение не дало бы положительного эффекта? Опять нет. Пока в распоряжении ***Альфы*** нет никакой дополнительной информации о факторах как взаимосвязанной системе, она не может выдвинуть никакой далеко идущей гипотезы, и ей остается лишь бесконечно экспериментировать в надежде на случайную удачу. А стороннего эксперта как источник такой дополнительной информации мы исключили.

Описанная выше проблема наводит на естественную мысль, что сама ***Альфа*** должна научиться извлекать дополнительную информацию. Нашу машину спасет естественное предположение, что в любой игре факторы, на нее влияющие, не равноценны, хотя бы в силу того, что есть ситуация под названием «Победа» или «Завершение игры». В шахматах мы знаем, что победа очень часто, почти всегда сопровождается приобретением материального перевеса. Это означает, что материальный фактор имеет доминирующее значение, и добавление веса ему до определенного осмысленного значения всегда хорошо.

Таким образом, исходным знанием ***Альфы*** о внешнем мире, каковым для нее является игра, есть знание о существовании доминирующих факторов. Причем это знание более высокого порядка, в сравнении со знанием качеств оценочной функции. Это своего рода фундаментальный закон, позволяющий разумно выстроить прикладную деятельность. Конечно, было бы интересно построить такую ***Альфу***, которая могла бы на основании опыта обнаружить этот закон. В самом общем смысле искусственный интеллект было бы интересно рассматривать как систему, способную только к методу проб и ошибок, одной интеллектуальной мета-операции, позволяющей открывать законы, улучшающие мышление, но это качественно иная задача. Сейчас мы предположим, что ***Альфа*** имеет генетически заложенное знание о существовании доминирующего фактора.

Это знание даст ***Альфе*** очень многое. Процедура поиска доминанты опирается на то, что его преобладающая величина дает возможность ***Альфе*** быть успешной против ***Беты*** при любой комбинации всех других факторов, достаточно лишь, чтобы доминирующий фактор имел значительную долю в весе оценочной функции. Что значит «значительная доля» – вопрос эксперимента, решаемый для каждой интеллектуальной задачи отдельно. Например, можно предположить, что его значение должно обеспечивать больше половины веса оценки, или вес доминанты должен быть выше веса любого другого фактора. Думаю, очень трудно дать здесь какие-то конкретные рекомендации о технике расчета «значительной доли» без знания конкретики задачи.

Метод корректировки оценки с опорой на доминанту

Выявление доминирующего фактора и расчет его оптимального веса в оценочной функции позволяют выстроить стратегию изменения весов оценки. Опираясь на грубую оценочную функцию, учитывающую только доминирующий фактор, можно выстроить приемлемую игровую стратегию и заняться подстройкой остальных факторов.

Вернемся немного назад. В чем была наша проблема? Для того чтобы выяснить, удачно или нет изменен вес того или иного фактора, надо провести много игр, и именно потребность в большом количестве экспериментов нас не устраивает. Известная доминанта позволяет выполнять корректировку в пределах одной партии. Введем понятие глубокого и короткого анализа позиции. Назовем коротким

анализом расчет полной оценки (с учетом всех факторов, в нее входящих) на небольшую глубину. Глубоким анализом будем называть расчет минимаксной оценки на значительное количество ходов, но только с учетом доминанты. Зная два результата, ***Альфа*** может сделать более точный вывод, выполнив следующую процедуру.

1. Первым шагом выдвигается набор гипотез: уменьшение или увеличение некоторого выбранного фактора улучшает качество игры. Таких гипотез можно выдвинуть сразу много.
2. Каждая гипотеза проверяется коротким анализом, по результатам которого часть построенного множества будет отброшена. Но для некоторых из них оценочная функция даст улучшенное значение. Выберем из них некоторое подмножество с оценкой, близкой к среднему значению всех полученных оценок.
3. Оставшиеся гипотезы проверяются минимаксной процедурой глубокого анализа, в котором участвует уже только доминанта. Если глубокий анализ подтверждает гипотезу, это резко повышает ценность гипотезы, так как доминирующий фактор, по большому счету, совпадает с целью игры.

Конечно, еще необходимо заметить, что гипотеза о более высоком весе фактора, подтвержденная в одной позиции, должна быть закреплена успешной проверкой и в последующих позициях.

Изменение состава оценочной функции

Выдвижение гипотез об изменении весов факторов – только полдела. Вторая половина – это изменение списка факторов. По большому счету, первичная гипотеза о составе исходной оценки может быть какой угодно. Можно первый набор факторов взять что называется с потолка. Но затем необходима какая-то рациональная стратегия. Строить ее будем, исходя из соображений здравого смысла, который нам подсказывает, что фактор, вносящий в оценку большое значение, безусловно, должен быть учтен. Фактор, чье влияние на ситуацию минимально, может быть отброшен. Это в отношении факторов, учитываемых в оценке. В отношении резерва можно выдвинуть следующее утверждение: если фактор только что был переведен из действующей оценки в резерв, возвращать его в оценку можно только после того, как оценка существенно изменится (а значит, может измениться и воздействие резервного фактора на игру).

Таким образом, можно выстроить следующий алгоритм в отношении факторного состава оценки.

1. Рассчитаем значимость каждого работающего фактора как среднее значение величин, вносимых фактором в оценку на протяжении ряда ходов.
2. Упорядочим факторы по значимости.
3. Фактор, находящийся внизу полученного списка, объявляется кандидатом на вылет.
4. Если значимость кандидата на вылет становится меньше некоей установленной критической величины, то он переходит в конец резервного списка.
5. А в оценку с нулевым весом переводится фактор с вершины резервного списка.

Доказывать разумность такой стратегии нет необходимости. Но обязательно укажем на слабое место наших рассуждений, которое является слабым местом вообще всех рассуждений о технологиях, построенных на оценке и дереве перебора. Впрочем, эту слабость можно отнести на счет некоторой поверхностности книги, которую вы читаете, так как ее главная цель – популяризация идей искусственного интеллекта, а популяризация всегда нуждается в упрощении и игнорировании деталей. Итак, что это за слабое место?

Заметьте, что очень часто приходится вводить в рассмотрение какие-то константы, определение значений которых само по себе представляет заметную проблему. Например, в алгоритме выше для вылета фактора из оценочной функции необходимо значение критической константы. Эта константа вряд ли является незыблемой на все времена. Разумно предположить, что она нуждается в непрерывной корректировке. Но, как уже было сказано, для того чтобы получить результат за ограниченное время, необходимо чем-то жертвовать. Например, мне необходимо сделать эту книгу читабельной, это само по себе ограничивает круг проблем, детальный анализ которых можно выполнить. Для программиста, пишущего игровую программу, нет смысла ставить задачу учета всех нюансов, если для этого потребуется слишком много времени. Нам всем приходится в своих задачах искать тонкую грань между качеством работы и ее результативностью.

Несколько идей общего характера. Короткий опыт

Выше были рассмотрены основные моменты технологии обучения машины, принимающей решение на основе оценочной функции. Те-

перь можно подумать и об общих вопросах машинной обучаемости. Прежде всего надо решить, чем отличается обученная программа от необученной. Что значит научиться?

Начнем с того, что искусственный интеллект, действуя во внешней среде, ставит перед собой определенную задачу, или эту задачу перед ним ставит кто-то другой, что не суть важно, но задача должна быть. Задачи могут самые различные, например – выжить в агрессивной среде, управлять производственным процессом в изменяющихся условиях и т. д. Нет задачи – нет и разумного действия. После того как задача поставлена, необходим критерий, позволяющий определить, насколько эффективно задача решается. Критерий должен быть построен с учетом поставленной цели.

Успешность производственного процесса можно измерять размером прибыли, успешность выживания можно измерять размером понесенного ущерба. И уже на этих двух примерах видно, что существуют критерии качественно разного порядка. Действия в агрессивной среде можно оценить по кратковременному эффекту. Производственный процесс от начала до получения реальных денег требует много времени и больших усилий. Есть цели с коротким шагом принятия решения (*сделали действие – получили понятный положительный* или отрицательный эффект). И есть длинные цели, для достижения которых необходима длинная цепочка решений, каждое из которых может быть верным, может быть ошибочным, но оценить реальный эффект можно только в конце. Ясно, что второй класс целей намного сложнее, поэтому разумно попытаться провести декомпозицию (разбиение на промежуточные цели) исходной задачи на более простые, в решении которых допустимо сиюминутное реагирование.

В качестве примера рассмотрим автомат, которому дано задание побродить по другой планете и дойти из некоторого пункта А в пункт В. Естественно, мы предполагаем набор возможных угроз, которые распознаются автоматом. Например, он может понять, что еще немного, и он перевернется, для чего ему достаточно вставить простейший прибор, определяющий угол отклонения от вертикальной оси. Он может понять, что еще немного, и он расплавится. Эта техническая проблема также легко решается обыкновенным термометром. То есть автомат имеет ряд приборов, позволяющих определить критические ситуации.

Еще необходимо добавить ему распознающее устройство. Этот распознаватель должен уметь рассматривать близлежащую окрестность и сравнивать новую информацию с хранящимися в памяти

образами ранее встречавшихся ландшафтов. Распознающее устройство способно выделять в ландшафте объекты, и если увиденного нет в памяти, то давать новому объекту имя и запоминать его. Объекты, конечно же, будут обладать определенной уникальностью. Например, асфальтированное шоссе (если таковое окажется на гипотетической планете) существенно отличается, скажем, от отвесной скалы или кратера действующего вулкана. Конечно, автомат, обнаружив шоссе или скалу, не будет знать, что это такое в нашем, человеческом понимании. Он будет в состоянии только понять, что он встретил объект, которого раньше не было, объект выглядит так-то, и имя ему отныне такое-то.

Цель интеллектуальной системы – найти связь между факторами угрозы и шаблонами встречающихся ландшафтов или возможных препятствий. Короткая задача такого автомата заключается в принятии решения о направлении движения при встрече известного препятствия и накоплении опыта при встрече неизвестного. Вот пара примеров возможных действий автомата.

Пример первый. Допустим, на пути встречено препятствие, которое с нашей точки зрения называется «непреодолимая отвесная скала». Для автомата это большая новость, и он даст препятствию имя, например ***A1***. Затем он попытается пройти через ***A1***. Этого не получится. Тогда в его памяти появится полезный опыт, и когда он в следующий раз встретит ***A1***, то уже не будет тратить время на попытку его преодоления, так как запомнит, что в предыдущей попытке он чуть было не перевернулся.

Пример второй. Автомат встретит возвышенность и даст ей имя ***A2***. Допустим, возвышенность не очень высока, и получилось ее преодолеть, не перевернувшись. Тогда, встретив ***A2*** второй раз, он попытается ее пройти. Допустим, второй раз попытка будет неудачной. Тогда в опыте машины зафиксируется результат, который мы, люди, называем «50 на 50». Третий раз автомат примет решение о штурме возвышенности с вероятностью 0,5. В случае удачи коэффициент успешности вырастет, в случае неудачи уменьшится.

Таким образом, автомат не выстраивает какой-либо долгосрочной стратегии поведения. Но, справляясь с локальными задачами, он тем не менее делает все возможное, чтобы приблизиться к цели. А при условии, что достичь пункта *B*, может быть, и нельзя, от автомата не стоит требовать большего. Его интеллект будет эффективен при условии достаточного набора устройств, анализирующих текущее состояние автомата и распознающих на этой базе потенциальные угрозы.

Несколько идей общего характера. Фреймы

Общее понятие термина «фрейм» – остов, набор главных характеристик, эскизная картинка. Например, мы могли бы составить фрейм человеческого лица, перечислив его элементы и дав каждому наиболее характерное описание. А можно просто нарисовать глаза, рот, нос, уши и т. д. в обычном, человеческом исполнении. Этого было бы достаточно для определения любого лица именно как человеческого. Такого рода фрейм в виде наиболее общего рисунка можно создать для любого визуально наблюдаемого объекта. Для сложного понятия фреймом может быть его определение. Для геометрической фигуры – идеальная фигура. Например, фреймом треугольника можно назвать фигуру, для которой не важны толщина линий, их цвет. Можно даже создать простой алгоритм конструирования фрейма. А именно возьмем конкретный объект, относящийся к данному классу (человеческое лицо, треугольник, уравнение, шоссейная дорога), возьмем этот объект, так как он дан в ощущениях. Затем разделим его на составные части и присущие ему свойства и начнем убирать их из общего набора. Тот минимальный набор, который сторонний эксперт еще может опознать как исходный объект, и будет фреймом.

Понятие исключительно полезно для создания эффективных механизмов памяти. Не берусь утверждать, что фреймовая технология лежит в основе функционирования нашей памяти, хотя и такая гипотеза имеет место, но, во всяком случае, фреймы дают возможность хранить огромные объемы информации. Возьмем, к примеру, человеческое лицо. Оно обладает огромным зарядом уникальности. Встретить двух совершенно одинаковых людей (не близнецов) практически невозможно. Полный человеческий портрет характеризуется большим количеством деталей, каждая из которых вносит что-то свое. А теперь вопрос. Можете ли вы ясно, в мельчайших подробностях представить в памяти лицо хорошо знакомого вам человека, даже того, кого вы видите каждый день? Думается, почти любой человек даст на этот вопрос отрицательный ответ. С другой стороны, увидев этого человека, вы не сомневаетесь, что это он, хотя, по сути, вы его и не помните. И даже более того, мы можем запоминать случайные лица, где-то раз увиденные, и, встретив его еще раз, даже спустя продолжительное время, мы готовы ручаться, что уже видели его.

Оба этих факта: в деталях восстановить не можем, но, увидев, не сомневаемся, – говорят о том, что в нашей памяти хранится некий

обрезанный образ, достаточный для вспоминания. Кстати, и с точки зрения эффективности работы памяти фрейма достаточно. Ведь память нам необходима для распознавания объекта и соотнесения его с какими-то знаниями о нем. Еще раз повторюсь, вышесказанное – не описание механизма работы человеческой памяти, а лишь идея, как ее возможно сконструировать.

Фреймы памяти помогают понять, каким образом некое мыслительное действие интерполируется на еще не встречавшуюся ситуацию, но имеющую аналог. Все очень просто. Будем считать два объекта одинаковыми с точностью до фрейма, если процедура построения фрейма для каждого из них дает один и тот же результат. Если же набор действий привязан к фрейму, то таким образом мы и обобщим этот набор действий на много объектов, имея опыт работы только с одним из них.

Конечно, это примитивная формулировка. Есть одна существенная проблема. Предположим, два объекта: ручей и горная река – имеют один фрейм, который мы условно назовем водная преграда. Лесной зверь (косуля), получив опыт прыжка через ручей, может обобщить его и на горную реку, что приведет к неудаче. Конечно, можно сказать, что падение в реку для лесного жителя не трагедия, а повод скорректировать набор фреймов. Но это плохо. Во-первых, такой выход из положения понижает ценность идеи, так как количество фреймов начнет резко расти, а мы их придумали для уменьшения требуемого объема памяти. Во-вторых, хотелось бы получать информацию до совершения опыта, так как неудача может привести к ущербу или даже летальному исходу, что не всегда приемлемо. Для комаров гибель нескольких миллионов особей не значит ничего, для пчел потеря десятков работников уже существенна. А для высокоорганизованных существ есть смысл бороться за каждую особь. Вернемся к той же косуле. Падение в реку не обязательно приведет к ее моментальной гибели, но она может сломать ногу, что фактически означает скорую гибель от хищника. В общем, есть смысл поискать другой путь.

Поэксплуатируем еще немного идею, использованную ранее. Я имею в виду идею доминантного признака. Во фрейме не все компоненты одинаково значимы для процедуры принятия решения. Например, если стоит задача выбора средства передвижения от пункта *A* до пункта *B*, то главным образом нужно знать расстояние до пункта назначения. Если эта величина в пределах пары сотен метров, то вопрос выбора вряд ли имеет смысл. Такое расстояние можно пройти пешком. Если расстояние в несколько сотен или тысяч километров, то этот фактор более чем существен. Если необходимо сложить пару

больших чисел, то сгодится и калькулятор, если же речь идет о сотнях или тысячах чисел, то разумнее взять в качестве инструмента компьютер и сделанную специально под задачу программу.

Теперь можно поточнее сформулировать идею доминантного признака. Есть объект – речка, есть субъект – косуля, желающая через речку перепрыгнуть и уже имеющая опыт прыжков через ручей. Есть фрейм, описывающий оба препятствия как водную преграду с доминантным признаком – ширина. Почему ширина – доминанта? Да потому, что в ней увязывается исходная задача косули с ее способностью к действию – прыжку определенной длины. Насколько далеко возможно прыгнуть, косуля осведомлена без привязки к заданному фрейму. И все, что ей сейчас остается, – это сопоставить свою способность прыгать с величиной доминантного признака, а значит, отпадает необходимость в приобретении нового опыта и составлении нового фрейма под названием «Горная река».

Конечно, это самая общая схема, но она легко поддается улучшению. Посмотрим, как это можно сделать на примере той же косули. Введем во фрейм «Преграда» доминантный признак под названием «Опасность». Его значение вполне может изменить поведение косули. Если косуля убегает от другого фрейма под названием «Хищник», в котором уровень опасности поднят до смертельного, а река обещает лишь возможные, хотя и крупные, неприятности, то прыгнуть стоит, даже если доминанта «Ширина» говорит о том, что допрыгнуть, возможно, не получится. Вот так, добавляя доминантные признаки, выстраивая их по приоритетам, сопоставляя их с возможностью выполнять те или иные действия, можно обучить систему выполнять реакции высокой степени сложности.

Несколько идей общего характера. Причинно-следственные связи

Термин «причинно-следственная связь» отражает важную особенность окружающего мира. В нашей вселенной можно выделить пары событий, таких, что одно из них, называемое причиной, неизбежно порождает другое, называемое следствием. Удар молнии в атмосфере неизбежно порождает гром, увеличение объема газа при постоянной температуре порождает уменьшение давления, рост скорости тела при его постоянной массе ведет к увеличению кинетической энергии (энергии движения) и т. д.

Иногда причинно-следственные пары не обладают свойством неизбежности. Например, попадание камня в оконное стекло не обязательно приведет к тому, что стекло разобьется, однако такую причинно-следственную связь мы принимаем. Если мы видим большую синюю тучу, то, видимо, будет дождь, но есть вероятность, что туча нас просто попугает. Пара событие–следствие может иметь статус неизбежности, но может быть и вероятностной. У описанных примеров причинно-следственных пар есть два общих момента. Во-первых, причина является причиной на том основании, что она запускает процесс, приводящий к следствию. Во-вторых, между причиной и следствием всегда есть разбег по времени. Не всегда его можно наблюдать визуально, но он всегда есть. Например, между включением фонарика и светом от него наблюдаемой временной задержки нет, но это только по причине очень большой скорости процесса. На самом же деле разница по времени есть всегда, все определяется только техникой наблюдения.

И вот эти две вещи: разбег по времени между причиной и следствием и существование связи – дают разуму инструмент, обеспечивающий ему возможность предсказывать события. Все, что нужно, – это максимально широкое наблюдение и увязка наблюдаемого со временем. Если замечено, что за событием *A* следует событие *B*, то появляется причина создать пару (*A*, *B*), которая будет усиливаться каждым новым положительным наблюдением, переходя из категории гипотезы в категорию закона, и ослабляться с каждым отрицательным наблюдением (*A* произошло, а *B* не последовало).

Здесь, конечно, тоже возникает масса технических вопросов. Например, какую разницу во времени считать достаточной. Например, приведет ли к засухе отсутствие дождей в течение недели? Видимо, нет. Можно ли ожидать неурожая картофеля через несколько месяцев, если сейчас ботва картофеля сильно поражена фитофторой? Видимо, да. Ответить на эти вопросы можно, используя доминантные признаки. Например, масштаб действия, его высокая энергетика вполне могут быть причиной быстрого результата. Большой разлив нефти на маленьком озере приведет к быстрой экологической катастрофе, а накопление знания по математике в течение урока потребует нескольких лет для развития математической культуры.

В общем, как и все инструменты развития интеллекта, этот инструмент требует решения комплекса вопросов, но его сила в том, что, используя базовую возможность наблюдения, можно выстраивать не только причинно-следственные пары, но и целые цепочки. Если из *A* следует *B*, а из *B* следует *C*, то из *A* следует *C* и т. д., сколь угодно

длинно. А если еще раз вспомнить, что причинно-следственные связи формируются во времени, то умение их видеть позволяет предсказывать события, а значит, учиться выстраивать стратегию поведения.

В заключение

Обучаемость состоит из трех важных вещей. Во-первых, обучаться значит увеличивать свою базу знаний об окружающем мире. Во-вторых, обучаться значит приобретать новые методы мышления. И в-третьих, обучаться значит знать, что ты знаешь и умеешь, а что еще недоступно. Последний пункт называется рефлексией, и я полагаю, что здесь прячется решение проблемы сознания – «Я не только мыслю, но я еще и осознаю, что я мыслю». О рефлексии и феномене сознания мы в этой книге еще поговорим. А сейчас несколько замечаний о расширении методологии мышления.

Интеллект способен развивать свою базу методов. Есть методы, выводимые из других. Наверное, можно создать алгоритм, по которому метод обобщения возникает на базе метода проб и ошибок. Собирая статистику своих действий в похожих ситуациях, интеллект создает причинно-следственные пары (делал то-то, получилось так-то). Похожие пары объединяются в группу, характеризующуюся общим признаком, и затем любая пара, обладающая этим признаком, автоматически попадает в эту же группу, и любое знание о группе автоматически переносится на новую пару без дополнительных экспериментов, мы это и называем обобщением.

Различных методов очень много, и большинство из них, видимо, сконструировано разумом. Но есть что-то, обладающее статусом базовых методов мышления. Таким базовым способом получения знания выглядит метод проб и ошибок. Наверное, можно выделить и другие изначальные вещи.

И здесь появляется ряд фундаментальных вопросов. Во-первых, можно ли выделить полную систему базовых методов, на которых строится все мышление? Во-вторых, если такая конечная система существует, то обязательно ли она одна? Если она не одна, то это означает возможность существования не просто других разумных существ, а существование других типов разума. И в-третьих, существует ли обобщенный механизм конструирования новых мыслительных методов? Все это крайне важные вопросы, требующие своего разрешения.

Глава 4

Сетевая архитектура

Искусственный интеллект – это в какой-то степени чистая математическая конструкция – набор алгоритмов, эвристических или каких-то иных, но алгоритмов. А любые алгоритмы нуждаются в устройстве для своего исполнения. И вопрос устройства тем более небезразличен для ИИ в силу необычности самой концепции интеллекта, уж больно он не похож на машину в нашем общечеловеческом понимании.

Машину, реализующую искусственный интеллект, можно построить сразу с жесткой структурой, отнеся вопросы гибкости и способности перенастраиваться с задачи на задачу на программное обеспечение. Однако факт существования мыслящего мозга показывает, что такой подход не единственно возможный. Наш мозг не появляется на свет знающим и умеющим все, что можно, это, во-первых, а во-вторых, его системотехника радикально отличается от компьютерной.

Человеческий мозг на нижнем уровне своей организации состоит из одинаковых элементов – нейронов. Один нейрон от другого не отличается, их конструкция очень проста, а функционал довольно примитивен. И все структуры мозга образуются из них в процессе развития. Несколько утрировано можно сказать так: изначально есть неорганизованная масса нейронов, которая, отвечая на внешние потоки информации, начинает процесс самоорганизации, постепенно выстраивая все необходимые структуры.

Выгода такой постановки дела очевидна. Элементы нижнего уровня – нейроны имеют очень простую структуру, а значит, их легко реализовать (если мы, конечно, поймем, какая структура нейронов необходима). Сборка любых частей машины, решающих различные задачи и при этом состоящих из одних и тех же компонентов, – мечта

любого производства. Кроме того, такая конструкторская идея открывает совершенно потрясающие возможности для ремонта. Если есть некоторое количество свободных нейронов, то ремонт мыслящей машины состоит из двух операций: отключения поломавшихся нейронов и подключения запасных. Выгоды очевидны, но и проблем встает немало. Даже поверхностный анализ дает очень сложные вопросы:

1. Очевидно, что простота нейрона должна включать тот минимум, который позволит объединившимся нейронам создать мыслящую машину. А как определить этот минимум?
2. Что может представлять собой механизм самоорганизации? Каким образом поток внешней информации заставляет нейроны собираться в какие-то конструкции?
3. Каким образом из простого функционала можно получить сложный? Конечно, нам известен прецедент человеческой электроники, когда базовый элемент – транзистор – участвует в создании совершенно различных устройств. Но все же транзистор – не единственный компонент элементной базы.

Вопросов много, но тем не менее в целом идея нейронной сети выглядит очень интересно, ясно, что если удастся дать хорошие ответы на эти и многие другие вопросы, то мы получим очень серьезную технологию не только для построения искусственного интеллекта, но и техники вообще. Фантасты, кстати, на такую возможность указали достаточно давно, создав литературные произведения, в которых примитивные организмы, объединяясь в коллектив, приобретают не просто грубую силу, а качественно иные возможности. Примеры такого рода мы можем найти и в нашей земной природе. Пчелиная семья обладает возможностями, не имеющимися у одиночной пчелы, играющей в улье роль нейрона. У пчел даже есть некий символьный язык, благодаря которому они в состоянии передавать друг другу информацию о местонахождении цветов. Еще один интересный пример – это колонии муравьев и термитов. И муравьи, и термиты не только умеют самоорганизовываться на решение рабочих задач. Как утверждают энтомологи, и муравьи, и термиты за счет выбора рациона кормления личинок умеют выращивать специализированных особей: более мелких, рабочих и оснащенных мощной головой солдат. Каким образом на это способны существа, не обладающие даже хорошо развитой нервной системой, – очень большой вопрос. Возможно, ответ на него будет таким же, как и на вопрос, каким образом специально организованные сообщества нейронов человеческого мозга могут рассчитать параметры реактивного самолета.

Но вернемся к нашей частной задаче – построения интеллекта. Любая система обработки информации, а интеллект можно рассматривать и с такой стороны, нуждается в высокой производительности. Способность быстро реагировать упирается в способность быстро обрабатывать информационные потоки. Однако до сих пор многие задачи, которые человек решает практически мгновенно, для современных компьютеров оказываются неимоверно сложными. Такова, например, задача распознавания знакомого лица в толпе людей – задача сложная математически, требующая огромных вычислительных ресурсов, не вызывает затруднений для человека. Почему так?

Ответ на этот частный вопрос получен достаточно давно, и он заключается в простой фразе «параллельные вычисления». Что это такое, можно пояснить на очень простом примере алгебраических квадратных уравнений. Мы помним, что общий вид квадратного уравнения таков:

$$ax^2 + bx + c = 0.$$

Для его решения необходимо вычислить величину, называемую дискриминантом, по формуле:

$$D = b^2 - 4ac.$$

Если полученное значение больше нуля, то корней будет два, расчет которых можно вести независимо друг от друга. Но вот в чем беда, в однопроцессорном компьютере расчеты ведутся последовательно. Нельзя вычислять оба корня одновременно. Но если компьютер будет иметь два вычислительных ядра, то на этапе расчета корней время на вычисления можно сократить в два раза, передав счет первого корня одному ядру, а счет второго – другому.

Это, конечно, примитивный пример, но суть дела он показывает верно. Задач, сводящихся к одному линейному вычислительному процессу, очень мало. Можно даже утверждать, что любая достаточно сложная задача допускает возможность параллельных вычислений. Поэтому уже на заре развития компьютерной техники возникла идея многопроцессорных вычислительных систем. Рост производительности одного процессора упирается в технические ограничения, в то время как увеличивать количество процессоров можно практически бесконечно. Нейронная сеть в каком-то смысле является предельным вариантом развития многопроцессорной системы, в которой процессоров уже миллионы или даже миллиарды, каждый из них умеет мало, а вычислительная сила достигается за счет распараллеливания

вычислений и сосредоточения на одной задаче большого количества таких маленьких вычислителей.

Нейрон

Прежде всего заметим, что дальше разговор пойдет о моделировании нейронных сетей на базе существующей вычислительной техники. Архитектурно мы не отходим от идеи фон Неймана. Речь не идет о создании машин с новой архитектурой, пока нейронные сети – это специально организованные программы для классического компьютера. Конечно, какие-то работы в направлении создания совершенно новых машин идут, но мне, автору этой книги, неизвестны какие-либо яркие прорывные результаты в этой области. И полагаю, и не я только, но и более компетентные в этой области люди, что прорыв будет возможен только с созданием принципиально новой элементной базы.

Итак, мы с вами договорились, что дальше речь пойдет только о математической модели, и в первую очередь определим понятие нейрона. В дальнейшем мы не будем различать искусственный нейрон и естественный (клетка высокоорганизованной нервной системы), будем говорить просто о нейроне.

Единственная функция нейрона заключается в преобразовании сигнала. Есть входной сигнал, который он получает либо от входного устройства, либо от другого нейрона, затем он выполняет с полученным сигналом какие-то действия и выдает его как выходной, например, другому нейрону. И вот здесь кое-что необходимо уточнить.

В любой информационной системе всегда есть шум. Его природа может быть разной. Если система занимается передачей электрических импульсов, то всегда будет некоторое количество паразитных сигналов, не несущих в себе никакой информации и появившихся либо в результате внешнего воздействия (внешние электрические поля и т. д.), либо в результате неуправляемых внутренних процессов. Это означает, что если нейрон станет реагировать на любой входящий сигнал, то он начнет воспроизводить информационный шум, в котором со временем вся система просто утонет.

Заметим, что описанная проблема не новость, она известна и современным цифровым технологиям. Компьютер построен на устройствах, моделирующих триггер – элемент, могущий находиться в двух состояниях: 1 – включен и 0 – выключен. В технической реализации наличие электрического шума приведет к тому, что состояние «вы-

ключен» окажется невозможным в силу гарантированного минимального сигнала. Выход из положения заключается в том, чтобы состояние «выключен» определялось не нулевым сигналом, а меньшим некоего порогового значения, которое определяется так, чтобы порог был выше уровня шума. Ну а состояние «включен» мы определим сигналом выше порога.

Таким же образом можно определить и состояние активности нейрона. Активация означает готовность нейрона принять и обработать поступивший сигнал. Отличие от старого доброго триггера здесь в том, что для нейрона в общем виде определяется не один сигнал, а группа, по которой вычисляется суммарный, или средневзвешенный, сигнал, и если он оказывается выше порога, то нейрон активируется. Надо сказать, что в теории нейронных сетей сказано довольно много о виде таких функций, но наша задача – уяснение общих идей, поэтому углубляться в математические тонкости не будем.

Второе отличие нейрона от триггера еще более интересно. Если задача триггера – лишь изображать собой ноль или единицу, то есть он устройство, в каком-то смысле пассивное, то нейрон соединяет в себе обе функции, он занимается и хранением информации, и ее обработкой. Активизировавшись, нейрон включает функцию обработки сигнала, которая, в принципе, может быть простой. Обработанный сигнал поступает на выход и передается другим нейронам или выходным устройствам.

Активность нейрона может быть угнетена. Предположим, что средневзвешенный входящий сигнал вычисляется по формуле:

$$F = \sum_k w_k x_k,$$

где x_k – это уровень k-го сигнала, а w_k – его вес. Если у некоторого сигнала высокий уровень, но ему приписан отрицательный вес, то этот сигнал будет работать на подавление активности нейрона. То есть даже такая простая функция счета входного сигнала создает интересные возможности для управления активностью нейрона. А это, в свою очередь, означает, что поведение нейронной сети зависимо от конкретной схемы подключения нейронов внутри сети, или, как говорят специалисты, зависимо от топологии сети.

Обучаемость нейронной сети

Нейронные сети, кроме возможности распараллеливания вычислительных процессов, предоставляют качественно новую способность

к обучению. Их обучаемость зашита в самой архитектуре сетей, является их фундаментальным свойством. Сеть – это прежде всего нейроны с определенной функцией преобразования сигнала. Эта функция, очевидно, не изменяема в процессе функционирования сети. Можно придумать сети с группами нейронов, обладающих различными функциями преобразования, можно создать сеть из однотипных элементов. Но сложность сети не объясняет ее способности обучаться.

Что означает фраза «Человек обучен» с точки зрения структур головного мозга? Новые нейроны в результате успешного учебного процесса не появятся. Устройство отдельных нейронов останется прежним. Единственная возможность что-либо изменить – это управление структурой связи между отдельными нейронами. Наверное, идеальным системотехническим решением, эффективным и то же время экономным, были бы блуждающие связи, вроде валентных связей атомов в молекулах. Один и тот же атом может вступать в химическую связь с разными атомами, создавая при этом совершенно разные по своим свойствам вещества. Но это пока из области фантастики. Более реальная схема – это соединение отдельного нейрона большим количеством равнозначных связей с другими нейронами и функцией активации, способной некоторые связи делать активными, а некоторые подавлять. Фактически такая функция будет изменять структуру мыслящей сети, а значит, изменять ее свойства. В общем, обучаемость нейронной сети зашита именно в этой функции.

Для вычислительных сетей есть термин «обучение с учителем». Это означает, что существует входной сигнал, для которого известна правильная реакция. Система, получая сигнал, сравнивает свою реакцию с заданным ответом. Если реакция ответу не соответствует, то это повод для изменения внутри сети

Сейчас, на некоторое время, чтобы дать понимание идей сетевой архитектуры, например то, как сетевое устройство может обучаться решению задач, мы отойдем от нейрона в том смысле, как его понимает современная теория искусственного интеллекта, и будем использовать понятие «вычислительная ячейка». Ячейка умеет выполнять какие-то простые операции, умеет получать данные от других ячеек и, естественно, передавать результаты своей работы. Нейрон – это частный случай такой ячейки. Понятие нейронной сети мы пока заменим понятием «вычислительная сеть».

А теперь вернемся к нашему подзаголовку. Подстройка – это своего рода обучение с «учителем». Роль учителя играет входной сигнал (набор входных данных), для которого известен правильный выходной сигнал (результирующие данные). Для лучшего понимания разберем технику подстройки на примере. Пусть набор входных данных состоит из числовых значений переменных *A*, *B*, *C*. Результирующий сигнал представлен величиной *Q* с известным числовым значением. Повторимся, числовые значения входных и выходных данных известны. Допустим, вычислительная сеть выдала результат, совпадающий с контрольным сигналом, в этом случае подстройка системе не нужна.

Теперь предположим более интересную ситуацию. Выходные данные не соответствуют контрольному сигналу, то есть вычислительная сеть ошиблась. Положим, что в сети есть ячейки, увеличивающие значения выходных величин, и есть – уменьшающие. Если полученная величина *Q* больше исходного значения, то надо в ячейках, увеличивающих *Q*, повышать отрицательные веса. Такая политика начнет торможение ячеек, создающих ошибку, в результате можно ожидать, что величина *Q* начнет уменьшаться, а ошибка между контрольным значением и вычисленным сетью – сокращаться. Спустя некоторое количество учебных запусков сеть может прийти к значению, отличающемуся от контрольного, в допустимых пределах. Покажем на примере, как это возможно.

Сконструируем вычислительную сеть, вычисляющую сумму или разность двух положительных чисел (A и B), и покажем, как она может настраиваться на одну из доступных для нее задач. Построим сеть из трех ячеек: *P* (инициализирующая результат), *Q1* (прибавляющая 1 к результату), *Q2* (вычитающая 1 из результата). Подключим ячейку друг к другу следующим образом (см. рис. 4.1):

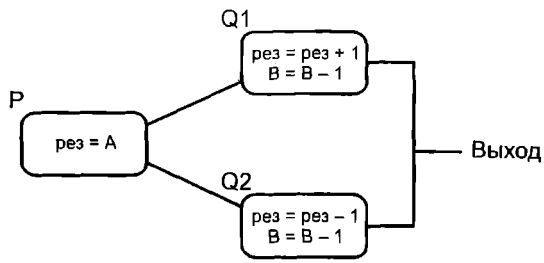

Рис. 4.1 ❖ Суммирующая сеть

Чтобы не загромождать схему, здесь указана только самая общая информация, а работать схема будет так: ячейка *P* получает числа A и B и вычисляет исходное значение величины **рез** (результат). Затем данные передаются ячейкам *Q1* и *Q2*. Они активируются безусловно, только фактом получения информации. *Q1* начинает наращивать величину «рез» и уменьшать величину B. Ячейка *Q2* выполняет обратную работу. Алгоритм работы ячеек таков:

```
Пока B не равно нулю, делать
      Изменять величину «рез»
Конец цикла
```

Пусть теперь сеть получила на вход набор данных (2, 3) и контрольное значение, равное 5. Такой контрольный сигнал будет отражать операцию сложения. *Q1* сможет закончить свою работу и выдать ответ, соответствующий контрольному значению.

Модифицируем сеть, введя функцию активации для Q1 и Q2. Оформим функцию максимально просто. Пусть она определяется одним числом w и ячейка активна, только если это число больше нуля. До начала работы инициализируем значение w единицей. И создадим дополнительный канал между *Q1* и *Q2*, по которому можно передать новое значение w (см. рис. 4.2).

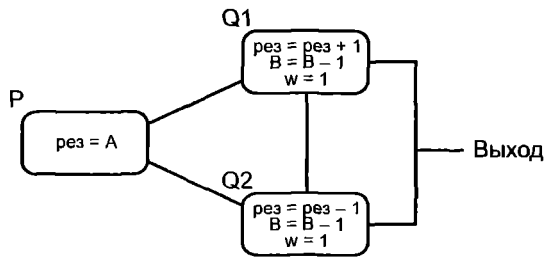

Рис. 4.2 ❖ Подстройка суммирующей сети

Внесем в работу ячеек *Q1* и *Q2* корректировку. Пусть теперь закончивший свою работу первым отошлет по каналу, связывающему его с партнером, отрицательное значение для w. Тогда наш контрольный сигнал (2, 3, 5) приведет к установлению постоянной активной цепочки *P – Q1 – Выход*. А значит, сеть настроится на выполнение сложений двух чисел, а *Q2* будет деактивирована значением, полученным от *Q1*. И можно быть уверенным, что последующие примеры на сложение также будут выполнены правильно. Перенастройка

сети на вычитание двух положительных чисел может быть произведена инициализацией функции активации и выполнением простейшего примера на вычитание, после которого сети можно дать большой пример.

Построенная нами вычислительная сеть, конечно, игрушечная и не пригодна для реальных задач. И ячейки слишком сложны (умеют выполнять законченный алгоритм счета), но это не более чем иллюстрация, цель которой – показать принцип работы сети, способной настраиваться на выполнение задачи, создавая устойчивые активные связи, которые, в принципе, потом можно разрушать и создавать новые, под новые задачи. При этом сама сеть не изменяется, возможности каждой ячейки не изменяются, модификации подлежат только условия взаимодействия ячеек друг с другом. И чем их больше, чем сеть обширнее, тем больше возможностей по созданию сложных конфигураций. И что очень интересно, активные конфигурации ячеек создаются не программистом, а актом настройки сети, ее обучением на задачу.

Специализированную сеть, умеющую настраиваться на задачи из ограниченного класса, придумывать не так уж сложно. Такие сети могут быть очень просты в реализации. Рассмотрим для примера задачу сортировки в самом общем ее понимании

Если дано множество чисел, то его можно отсортировать по возрастанию, можно по убыванию. А можно распределить числа по какой-то кривой (как, например, на рис. 4.3):

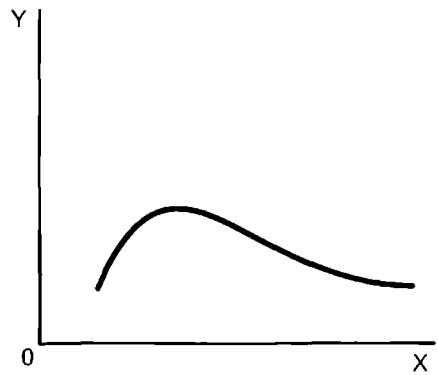

Рис. 4.3 ❖ Произвольная кривая

Всевозможных распределений множества чисел существует сколь угодно много. И алгоритмов сортировки тоже множество. Посмотрим, сможет ли вычислительная сеть справиться с общей задачей сортировки. Предпосылки к этому есть, и очень хорошие. Напомним, что простейший алгоритм сортировки по возрастанию (убыванию), именуемый пузырьком, за одну итерацию сравнивает и переставляет два рядом стоящих элемента. Ясно, что можно придумать линейную сеть, такую, что изначально в каждую ячейку загружается одно число, а затем ячейки начинают обмениваться друг с другом по какому-то правилу. Конечно, не слишком интересно специализированное правило. Хотелось бы придумать самую общую схему, позволяющую настраиваться на любое распределение.

Нам понадобится линейная сеть, способная запоминать шаблон и подгонять любой входной набор данных под этот шаблон. Для упрощения рассуждений положим, что количество ячеек в точности соответствует количеству сортируемых чисел. Иное составляет проблему технического, а не принципиального характера. Построим линейную сеть такого вида (рис. 4.4):

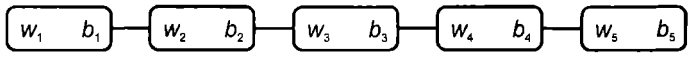

Рис. 4.4 ❖ Линейная сеть для сортировки

Конструктивно устроим так, чтобы ячейки регулярно устанавливали контакт в парах (образованных соседями). Как именно будет это происходить, в каком порядке, для нашей задачи роли не играет. Желательно лишь, чтобы интенсивность контактов между любыми двумя парами была примерно одинаковой. Цель контакта – обмен данными. Но если любая пара, установившая контакт, будет в обязательном порядке обмениваться данными, то толку с этого не будет, необходимо установить какие-то правила взаимодействия. Опишем некоторые желательные свойства нашей сети.

Характеристики w_k используем для хранения обучающего сигнала. Если сеть инициализирована, скажем, нулями, то первый сигнал она воспримет как обучающий и сохранит полученные данные в величинах w.

Введем следующее правило обмена данными. Есть пара ячеек, установившая контакт. Первая из них хранит характеристику w_1 и данное b_1. Вторая имеет характеристику w_2 и данное b_2. Обмен между ними будет осуществляться только в том случае, если выполняется условие:

$$\max(|w_1 - b_1|, |w_2 - b_2|) > \max(|w_1 - b_2|, |w_2 - b_1|).$$

Левая часть неравенства представляет собой максимальную ошибку исходного распределения. Правая часть – это максимальная ошибка после возможного обмена. Естественно, если максимальная ошибка уменьшается, то необходимо запустить процедуру обмена. Небольшой пример. Построим сеть из двух ячеек, моделирующую проверку неравенства A < B и обмен значениями между величинами A и B, в случае если неравенство оказывается истинным (см. рис. 4.5):

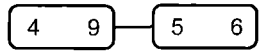

Рис. 4.5 ❖ Пример неравенства

Наша сеть была обучена на примере (4, 5), что видно по весовым значениям, а данные, полученные на обработку, – это (9, 6). Сейчас модуль максимальной ошибки 9 – 4 = 5, после возможного обмена максимальная ошибка окажется равной 9 – 5 = 4. То есть обмен желателен. Наше условие выглядит естественно, и можно ожидать, что оно будет работать на значительном количестве примеров. Но не факт, что везде. Попробуем на этом условии отработать сортировку по возрастанию. Пусть сеть обучена контрольным сигналом (1, 2, 3) и загружены данные для сортировки (5, 4, 1). Естественно ожидать в качестве результата последовательность (1, 4, 5). Посмотрим, получится ли. Исходная картинка такова (см. рис. 4.6):

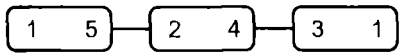

Рис. 4.6 ❖ Начало сортировки

Схему образования пар выберем самую простую. Первыми контактируют первая и вторая ячейки, затем вторая и третья, затем опять первая и вторая. Максимальная ошибка первой пары 5 – 1 = 4. В случае обмена максимальная ошибка уменьшается: 4 – 1 = 3 и 5 – 2 = 3. Ячейки обмениваются данными, и мы получаем такую картинку (см. рис. 4.7):

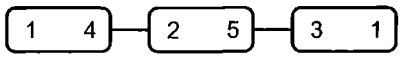

Рис. 4.7 ❖ Первый шаг сортировки

Теперь анализируем пару, составленную из второй и третьей ячеек. Имеем модуль максимальной ошибки 5 – 2 = 3. Если будет произведен обмен, то этот модуль составит 5 – 3 = 2. Обмен желателен, и следующий промежуточный результат таков (см. рис. 4.8):

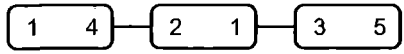

Рис. 4.8 ❖ Второй шаг сортировки

Возвращаемся к паре, составленной из первой и второй ячеек. В данный момент максимальная ошибка составляет 4 – 1 = 3. После возможного обмена модуль максимальной ошибки уменьшится: 4 – 2 = 2. Выполняем обмен и получаем следующий результат (см. рис. 4.9):

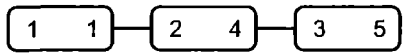

Рис. 4.9 ❖ Последний шаг сортировки

На этом шаге процесс, очевидно, остановится, так как любая попытка обмена в дальнейшем будет приводить к увеличению максимальной ошибки, и в это же время мы видим, что сортировка завершена. В сконструированной сети фактически реализован алгоритм пузырька. Но благодаря архитектуре сети (если, конечно, она реализована не на компьютере с неймановской архитектурой) процесс очень сильно распараллелен. Если, к примеру, в сети 100 ячеек, то на ней могут создаваться 50 одновременно работающих пар, если мы только лишь немного усложним технику работы сети (в нашем примере за раз работает одна пара, но мы можем образовывать много пар одновременно). А это, в свою очередь, означает повышение производительности в 50 раз. И чем сортируемых чисел больше, тем выигрыш в производительности будет выше. И конечно же, такая система должна быть на порядок дешевле, нежели многопроцессорный компьютер, все-таки процессор – это очень сложная и дорогостоящая вещь в сравнении с примитивным устройством ячейки.

Наша сеть обладает довольно значительной степенью универсальности. Точно так же, как мы за один прием обучили ее выполнять сортировку по возрастанию, инициализировав веса нулями и послав контрольный сигнал, представляющий собой массив, упорядоченный по возрастанию, мы обучим ее сортировке по убыванию, отправив

в качестве контрольного сигнала убывающий массив. Если дать ей массив чисел, располагающихся по гауссиане или вдоль другой кривой, она, очевидно, с последующими массивами будет делать то же самое. Но, конечно, придуманное правило уменьшения максимальной ошибки подлежит доказательству, хотя и выглядит разумно. Однако если искать доказательство лень, правило можно принять как полезную эвристику.

Сеть нелинейной геометрии

Заметим, что пока придуманные сети имеют линейную архитектуру. Попробуем на простом примере поиска наибольшего числа в массиве построить сеть с более сложной конфигурацией. Для конструирования сети потребуются ячейки, умеющие хранить одно число, способные его передавать, получать и сравнивать. Такие минимальные навыки заложить в простую схему можно. Конечно, нужна некая выделенная ячейка, в которой в конце процесса окажется найденное наибольшее число. Схема должна работать быстрее обычного алгоритма перебора, учтем, что работа алгоритма сводится к сравнениям и обменам, количество которых сопоставимо с количеством элементов массива. Линейная сеть для такой задачи не подойдет по причине низкой производительности. Если выделенная ячейка, предназначенная для хранения результата, окажется на одном конце сети, а наибольшее число – на другом, то для того, чтобы дотащить наибольшее значение до конца сети, потребуется не меньше операций, чем в обычном алгоритме для машины фон Неймана. Построим сеть немного другой геометрии (см. рис. 4.10):

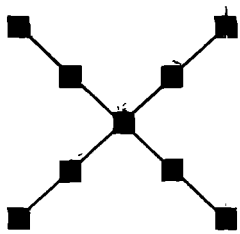

Рис. 4.10 ❖ Сеть для поиска наибольшего

Квадратики – это ячейки с описанными свойствами, соединяющие их линии – связи, по которым можно передавать данные. Придать сети нужные свойства несложно с помощью простой функции

активации. Ячейка может находиться в трех возможных состояниях: неактивна, настроена на передачу, настроена на прием. Пусть в начале процесса крайние настроены на передачу, остальные – на прием. Затем ячейка, принявшая число, перестраивается на передачу, передавшая число становится неактивной. Это условие запустит волну передач, направленную от краев к центру.

Небольшая проблема встанет с центральной ячейкой. После первого успешного сеанса приема она настроится на передачу, несмотря на то что ей необходимо попытаться принять еще три числа. Немного модифицируем функцию активации. Пусть блокирует ячейку значение функции активации, равное –1. 0 настраивает на передачу и число, большее нуля, – на прием. Еще договоримся, что любое действие уменьшает значение функции на 1. Тогда, если центральную ячейку инициализировать большим положительным числом, она долго будет настроена только на прием. Если крайние инициализировать нулями, то они после первой передачи перейдут в неактивное состояние. Остальных достаточно инициализировать 1. В этом случае первое действие (прием) переведет ячейку в состояние готовности к передаче, а второе (передача) переведет его в неактивное состояние. И наконец, в качестве собственного действия ячейки определим следующий простой алгоритм:

```
Если Собственное число меньше Полученного,
        То Собственное число = Полученному
```

Такая сеть сможет обработать массив в четыре раза быстрее. Для ее совершенствования можно увеличить количество ветвей, идущих к центральной ячейке, а можно сеть сделать многослойной, так чтобы центральная ячейка пучка была началом ветки пучка следующего уровня. Картинка ниже (рис. 4.11) иллюстрирует эту идею.

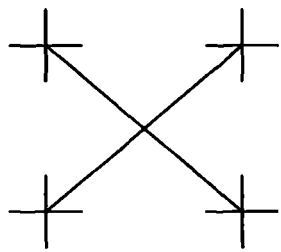

Рис. 4.11 ❖ Многослойная сеть для наибольшего

Это двухуровневая сеть. А так как количество уровней можно сделать сколь угодно большим, то и скорость поиска наибольшего можно увеличивать почти безгранично. Заметим также, что задача поиска наибольшего числа допускает естественное обобщение: дано множество объектов, описывающихся определенной характеристикой. Необходимо найти объект, для которого эта характеристика: принимает наибольшее значение, менее всего отличается от заданного, попадает в некоторый интервал. В общем, можно придумать много задач, для решения которых такая конструкция сети будет полезна. Конечно, функцию активации для иерархической сети придется усложнить, но это уже проблема технического, а не принципиального характера.

Персептрон Розенблатта

Проблема теории нейронных сетей, а может быть, ее достоинство, – в том, что она еще очень молода, как и вообще теория искусственного интеллекта. Поэтому в изложении разных авторов можно встретить нюансы понимания тех или иных концепций, разночтения если не в базовых терминах, то в их использовании. Еще один важный момент – необычность идей искусственного интеллекта, ограничивающая их понимание. Конечно, есть элита специалистов, для которых проблем восприятия идей ИИ не существует, однако эта книга не для них. Моя цель – дать представление основных идей в виде, доступном для неспециалиста, для чего, в частности, в этой главе допущены некоторые вольности. Например, мои вычислительные ячейки, использованные выше, довольно сильно отличаются по своим возможностям от математических моделей, предложенных отцами – основателями теории нейронных сетей. Но это общая проблема поиска золотой середины между популярным изложением и научным. Если первое рискует впасть в профанацию, то второе может оказаться слишком академичным и непонятным. Надеюсь, мы смогли уйти от обеих угроз, и базовая идея – конструирование устройства, состоящего из простых элементов, способных настраиваться на решение сложных задач, изменяя собственные свойства, вами понята. А если это так, то давайте посмотрим, что понимал под нейронной сетью американский нейрофизиолог Ф. Розенблатт. Персептрон (в некоторых транскрипциях «перцептрон») Розенблатта (рис. 4.12) – это модель нейронной сети, которая по его идее должна моделировать принципы работы человеческого мозга в деле распознавания изображений.

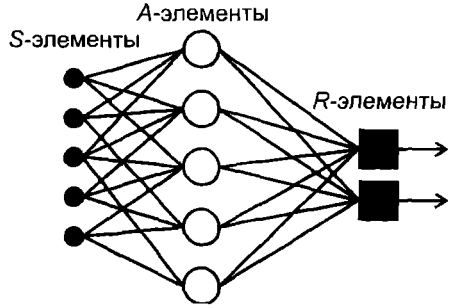

Рис. 4.12 ❖ Персептрон Розенблатта

Персептрон состоит из трех слоев устройств разной функциональности. Поэтому его разумно называть однослойной нейронной сетью, хотя в некоторых источниках о ней говорят как о трехслойной. Первый слой: *S*-элементы, получают входной сигнал и преобразуют его в вид, понимаемый *A*-элементами, собственно, и являющимися нейронами Розенблатта. *A*-элементы воспринимают сигнал и изменяют собственный вес по некоторому правилу. Умение изменять вес является основой обучаемости персептрона. И наконец, *R*-элементы выдают результат работы всего устройства. Сейчас попробуем разобраться, как это работает.

В режиме обучения схема сравнивает результат обработки получаемой картинки с эталоном. Получаемые извне картинки, будем их далее называть обучающими примерами, должны чем-то отличаться друг от друга, в противном случае обучения не будет. Эталон, конечно же, всегда один и тот же. Результат работы персептрона будет давать некоторую ошибку в сравнении с эталоном. Можно описать функцию, минимизирующую ошибку посредством изменения весов нейронов, определяющих вклад нейрона в результирующую картину. После достаточного количества исходных примеров можно ожидать, что персептрон научится выделять эталонную картинку из входного примера, после чего ему можно дать пример, не используемый в обучении.

Поясним, как это работает на геометрическом примере. Внизу три картинки: две учебные и эталон (см. рис. 4.13).

Все три рисунка содержат квадрат с эллипсами в вершинах. Функция, изменяющая активность нейронов, должна обеспечить небольшую корректировку формы и величины эллипсов и небольшую корректировку их местоположения. Если же персептрону подадут

примеры, в которых появится пятая фигура, то сигнал, за нее ответственный, должен быть максимально ослаблен. Если качественно понятно, что должно происходить, то посмотрим, каким образом деятельность персептрона Розенблатта обеспечена количественными соотношениями.

Рис. 4.13 ❖ Пример для распознавания

Введем обозначения: x^k – вектор входящего сигнала, y^k – вектор выходящего сигнала.

Шаг первый. Начальные веса всех нейронов в нулевой момент времени полагаются случайными:

$W(t = 0)$ = случайное число.

Шаг второй. Вычисляется вектор ошибки, каждая составляющая которого – это локальная ошибка в матрице входного сигнала. Вычисляется ошибка как разность между входным сигналом и эталонным:

$$\Delta_k = y^k_{\text{оригинал}} - y^k_{\text{эталон}}.$$

Шаг третий. Вектор весов (все веса, всех нейронов) для следующего момента времени изменяется по формуле:

$$W(t + \Delta t) = W(t) + \eta x^k (\Delta^k)^T.$$

Большая T за последней скобкой означает транспонирование вектора ошибки для перемножения его с вектором входного сигнала. С правилами и операциями векторной алгебры можно ознакомиться по любому учебнику векторной алгебры. Величина η обеспечивает скорость обучения персептрона, и поэтому она так и называется – темп обучения. Ее величина варьируется от 0 до 1. Принципиально функция работает очень просто. Если сигнал, выданный нейроном, недобирает до эталона, то он усиливается. Если нейрон выдал больше, чем нужно, то его активность подавляется.

Для иллюстрации работы персептрона Розенблата есть интересная геометрическая интерпретация. Заметим, что функция преобразования весов имеет линейный характер. Геометрически это означает, что персептрон пытается отделить полезную информацию от шума прямой линией. Это означает, что персептрон может распознать любой объект, отделимый прямой. Однако легко привести контрпример. На рисунке ниже два типа кружков. Черные и светлые. Допустим, что черные – это шум, а светлые представляют собой полезную информацию (см. рис. 4.14):

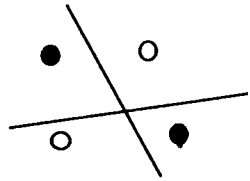

Рис. 4.14 ❖ Неустранимая проблема

Легко видеть, что нет такой прямой, которая отделила бы черные кружки от светлых. А значит, возможности линейного персептрона по распознаванию образов сильно ограничены. Впрочем, это означает лишь необходимость усложнения устройства. И направление усложнения вполне понятно. Выше было сказано, что персептрон Розенблатта однослойный, в том смысле что слой нейронов только один. Здесь зарыт колоссальный резерв усложнения, заключающийся в добавлении слоев и усложнении механизмов, передающих информацию между слоями.

Сети Кохонена

Теуво Калеви Кохонен – финский ученый, специалист в области искусственного интеллекта, предложил вариант архитектуры нейронных сетей, решающих задачу кластеризации объектов. Не перепутайте два различных понятия: классификация и кластеризация. Если первое из них означает анализ входных данных и отнесение объекта к какому-либо из уже известных классов, то второе означает отнесение объекта к одному из еще неизвестных классов. Звучит определение кластеризации несколько парадоксально, поэтому давайте разберемся с термином. Какой смысл относить объект к классу, который сам еще нуждается в определении?

А смысл есть, и достаточно солидный. Кластеризация, по большому счету, представляет собой начало любого исследования незнакомого окружения. Прежде всего, получив на анализ множество объектов, мы должны выделить из них похожие между собой, то есть, возможно, являющиеся модификацией одного объекта. Тем самым мы отделяем этот объект от других, принципиально от него отличных. Выполнив такую работу, мы уже можем дать объекту имя и начать исследовать его свойства.

Сеть Кохонена предполагает, что количество кластеров изначально известно. Это не такое уж сильное предположение. В каком-то смысле ошибиться в оценке их количества невозможно. Просто процесс кластеризации может провести свою работу более тонко или более грубо. Например, сеть может выделить в один кластер стулья, кресла, табуретки. Такой кластер вполне возможен, так как функционально эти три типа объектов похожи. Но при более тонком делении возможно образование не одного, а трех кластеров.

Разберемся, как сеть Кохонена может отделить стул от табуретки, или, в более общем виде, как она может разнести объекты *A* и *B* по разным кластерам. Договоримся, что для каждого объекта существует описание, представляющее собой упорядоченный набор признаков. Даже положим, что этот набор представляет собой множество чисел, или, как говорят математики, вектор (множество чисел, в котором роль играет не только значение, но и местоположение числа во множестве – его номер).

Договоримся, что описание любого объекта представляет собой одинаковые по длине наборы чисел, или, иначе говоря, векторы одной размерности. Это не слишком страшное ограничение. Так как нулевые значения не несут в себе никакой информации, то можно составить вектор, в котором некоторые величины будут нулями, то есть *добавить нулей до определенной длины вектора*. Такой вектор сеть воспринимает первым слоем нейронов, а значит, в первом слое количество нейронов равно размерности вектора описания объекта.

Второй слой состоит уже из любого количества нейронов, называемых линейными сумматорами. Сумматор называется линейным по причине принципа своей работы, он воспринимает входящий вектор и преобразует его в число следующей линейной функцией:

$$F = \sum_k w_k x_k,$$

где величины x_k – значения входящего вектора, а w_k – набор весов, свой, особенный для каждого нейрона. Такой линейный сумматор

выдает для каждого нейрона число. Конечно, в зависимости от выбора весов некоторые нейроны могут выдать одинаковое значение (не отличить стул от табуретки). Далее выбирается нейрон, выдавший наиболее сильный сигнал, а все остальные обнуляются. Такой метод формирования выходного сигнала обозначается определением «Победитель забирает все».

Из вышесказанного следует, что количество нейронов второго слоя равно количеству кластеров, на которые сеть может разделить входные объекты. Заметим также, что сеть Кохонена не нуждается в предварительном обучении с контрольным выходным сигналом. Графически сеть можно представить так (рис. 4.15):

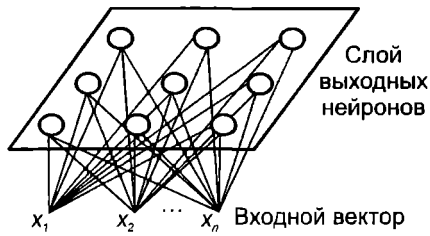

Рис. 4.15 ❖ Сеть Кохонена

Заметим, что конструкция обрабатывающих нейронов очень проста, вся смысловая нагрузка ложится на функции суммирования, на подбор весов. Но это уже отдельная математическая проблема, решаемая с учетом природы задачи. Задача кластеризации, заметим, легко преобразуется в задачу классификации, если о кластерах, на которые мы будем разделять объекты, что-то известно. Многослойные сети вполне могут использоваться для задач обобщения. При нескольких вторых слоях результат их работы можно интерпретировать как входной вектор для третьего слоя. Пусть, например, первый слой выделяет: стулья, табуретки, кресла, тарелки, блюдца, чашки. Сумматоры третьего слоя можно подобрать так, что первые три типа предметов будут выделены в кластер, который условно назовем мебель для сидения, а следующие три – в кластер посуда.

Многослойные сети Кохонена можно использовать для любого уровня обобщения, так как количество слоев может быть любым. Область использования сетей практически ничем не ограничена, точнее, ее функционал задается лишь нашей способностью представить входной набор данных вектором: числовым или символьным. Кстати,

числовое представление совсем не обязательно. Может быть, даже более интересно будет рассмотреть возможности сети на символьном представлении. Не таким уж невероятным выглядит и сумматор на символах. Как вариант рассмотрим следующее представление. Назовем базовым множеством сети множество символов $A = \{x_k\}$. Поставим каждому символу в соответствие числовой код по какому-либо правилу, например код может быть равен номеру символа в базовом множестве. Тогда работа сумматора ничем не отличается от описанной выше, сумматор символьной сети обрабатывает кодовые представления и находит их сумму.

Затем в базовом множестве ищется символ, чей код равен сумме. На тот случай, если сумма окажется больше максимально возможного кода, ее можно преобразовать к остатку от деления суммы на максимально возможный код. Возможных символьных функций можно придумывать очень много, вопрос лишь в том, будут ли они соответствовать какой-либо реальности. Но это вопрос, приложимый к любому математическому результату.

Звезды Гроссберга

Моделей нейронных сетей с различными свойствами, ориентированных на ту или иную задачу, довольно много, и для описания даже базовых идей, если делать это детально и обстоятельно, можно написать отдельную книгу. Здесь и сейчас задача создания энциклопедии не стоит, однако, как говорится, «бог любит троицу», поэтому завершим перечень примеров моделью, возникшей на заре теории, так называемыми «звездами Гроссберга». В каком-то смысле сети Кохонена можно считать обобщением идеи звезд. Обе конфигурации настраиваются на определенные образы, но если сеть Кохонена может быть настроена на N образов, то звезда Гроссберга – на один.

Звезда – это сеть со множеством входов и одним выходом. На вход подаются сигналы X_i, на выходе – один сигнал Y. Входные нейроны снабжены весами W_i. Выходной сигнал считается как средневзвешенная сумма. Формула такой суммы в тексте уже использовалась.

На вход звезды поступают немного отличающиеся сигналы. Отличающиеся в силу того, что они представляют собой описание различных экземпляров объекта, а отличаются они немного в силу того, что это разные экземпляры одного и того же объекта. Настройка звезды на образ объекта происходит подстройкой весов по формуле:

$$W_t = W_i + \alpha(X_i - W_i).$$

Здесь величина α – коэффициент обучаемости звезды, со значением от нуля до единицы. После нескольких сеансов звезда настраивается на свой объект, и если отключить возможность перестройки весов, то «свой объект» звезда будет распознавать достаточно уверенно, естественно, если сообщество входных нейронов было достаточно велико. Если включить режим обучения (опять разрешить звезде изменять веса), то она после нескольких сеансов перестроится на другой объект. Но, конечно, если звезда на входе будет получать образы различных объектов, то никакого обучения не получится.

В заключение

Мы рассмотрели примеры нейронных сетей на задачах распознавания объектов. И сегодня теория распознавания представляет собой главное поле применения сетей. Наверное, так будет не всегда. Человеческий мозг являет нам пример нейронной сети, обладающей универсальными навыками к самообучению. И будем надеяться, что человеческой науке с развитием элементной базы удастся повторить успех эволюции. Но сама по себе возможность создавать совершенные нейроны еще не будет означать появления совершенного искусственного интеллекта. В этом вопросе, пожалуй, более важную роль играют вопросы нейронной организации. Понять, каким образом группа нейронов может соответствовать интеллектуальной задаче, может быть еще более сложно, чем создать нейрон. Но возможности открываются действительно колоссальные. Если природа за миллиарды лет создала один тип нейронов, то мы, люди, уловив принцип, сможем создавать функционально разные нейроны, с заданными свойствами, скоростью обработки информации, намного превышающей возможности человеческого мозга. Но пока самые лучшие нейронные сети, созданные человеческим мозгом, даже близко не приближаются по возможностям к своему создателю.

Глава 5

Распознавание образов

Задача, заявленная в заголовке, намного шире визуального распознавания, хотя зачастую она воспринимается именно так. И на заре создания искусственного интеллекта речь шла именно об образах, получаемых от внешнего мира. Вообще, любое мышление начинается с информации. Мы что-то видим, слышим, осязаем, и первым шагом необходимо понять, что именно мы видим, слышим и осязаем, необходимо отделить существенную информацию от информационного шума. На рис. 5.1 – простой пример.

Рис. 5.1 ❖ Простой пример

Думаю, никто не будет спорить с утверждением, что здесь изображен треугольник, несмотря на то что сбоку от него есть еще какая-то черточка. Человек, обладающий даже минимальным математическим образованием, согласится, что черточка не несет в себе информации в сравнении с треугольником, то есть это шум. А вот этот рис. 5.2 уже имеет другую природу.

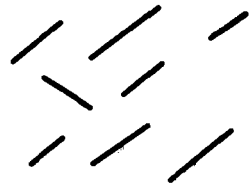

Рис. 5.2 ❖ Уже не столь простой пример

Вроде бы тоже черточки, но их сонаправленность несет в себе какое-то сообщение, и одна черточка, выпадающая из общего хора, уже может быть носителем информации. В принципе, ее можно объявить шумом, но можно сказать, что эта единственная черточка несет в себе больше информации, чем любая другая на этом рисунке, так как именно она утверждает, что сонаправленность не является жесткой закономерностью этого изображения.

Поэтому ответ на вопрос – что же мы видим на самом деле, не всегда однозначен. А есть еще и вторая часть проблемы распознавания. Необходимо дать имя тому, что ощущаешь, сопоставив его с уже известными вещами (*шаблонами*). Техника ответа на эти два вопроса:

1. Что мы видим, слышим, ощущаем?
2. Как назвать то, что дано нам в ощущениях?

– и является предметом теории распознавания. Возможна и более широкая трактовка задач распознавания. Начинается мышление не всегда с распознавания чувственных образов. Точно так же полноправным участником мыслительного процесса являются образы, данные не в ощущениях, а в самом мышлении. Вполне правомерен вопрос, как распознать некое утверждение. Например, формула

$$a^2 + b^2 = c^2$$

вполне может являться алгебраическим выражением теоремы Пифагора, но может оказаться, в зависимости от контекста, в котором она используется, целочисленным уравнением или уравнением окружности. Высказывание

Сумма внутренних углов любого треугольника равна 180°

для человека, не имеющего никаких математических знаний, не означает ничего, а для владеющего хотя бы начальными знаниями геометрии это – теорема о сумме углов треугольника. Вопрос, как интеллект распознает образные или речевые объекты и соотносит их

с названиями областей знания, теорем и т. д., очень интересен и намного сложнее, чем распознавание звуковых или зрительных образов, но сейчас мы займемся все же последними; теория образов, понимаемая именно так, достигла довольно существенных успехов, и это действительно интересно, каким образом научить машину отличать один объект от другого.

Выделение объекта из среды

Наиболее вероятна следующая ситуация: наблюдатель видит перед собой некое нагромождение точек, линий, поверхностей и т. д. В этом множестве, возможно, находится несколько объектов для распознавания, и, скорее всего, то, какие именно объекты там есть, изначально неизвестно. Вообще, надо сказать, контекст исследования, знание того, что мы ищем и как эта искомая вещь может выглядеть, очень сильно меняют ситуацию. Если известно, что перед наблюдателем текст на русском языке, то можно сказать, что известно очень много, перед нами набор символов кириллицы. Такое дополнительное знание дает возможность создавать специализированные и очень эффективные методы распознавания.

Чаще перед специализированным наблюдателем нет задачи распознавания неизвестного, класс объектов вполне определен – это горы, дома, автомобили, озера, буквы кириллицы, японские иероглифы, земноводные животные, камни-самоцветы и т. д. Но в любом случае, необходимо уметь выделять объект на картинке. Причем желательно уметь это делать без привязки к какому-то классу известных предметов. Другими словами, нужны общие методы отделения, использующие самые общие свойства сущности «объект», без конкретизации на какой-то класс.

Может показаться неожиданным, что можно выделить объект, не предполагая, что он из себя представляет, но это действительно так. Подумайте, каким образом вы приходите к выводу, что наблюдаете линию, как, например, на рис. 5.3:

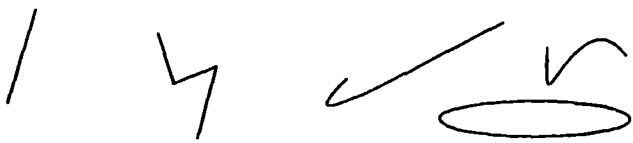

Рис. 5.3 ❖ Примеры линий

Здесь действительно нарисован набор линий – прямая, ломаная, кривые, есть одна замкнутая линия. Но если не рассматривать признаки формы, каковые являются признаками более высокого порядка, то это все линии. Как наблюдатель может прийти к выводу, что перед ним именно линии, и как он может отделить одну от другой?

Есть для этого очень простой критерий. Если две точки *А* и *Б* принадлежат одному объекту, то существует путь (линейное множество точек) от *А* к *Б*. Термин «путь» означает, что для любой точки этого линейного множества существуют две соседние точки, расстояние до которых находится в каких-то заданных пределах. Понятие заданного предела расстояния очень важно. Точки пути не обязаны прилегать вплотную друг к другу. Рисунок 5.4 иллюстрирует эту мысль.

Рис. 5.4 ❖ Три объекта на 12 точек

Даже интуитивно понимаемый критерий пространственной близости говорит о том, что перед нами не один объект из 12 точек, а три различных. А сейчас сформулируем названный критерий в хорошей алгебраической форме. Для упрощения ситуации положим, что наблюдаемое пространство двумерное. Тогда, согласно метрике Евклида (правило, описывающее способ вычисления расстояния между двумя точками на основе теоремы Евклида), расстояние вычисляется по формуле:

$$L = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

В этой формуле L – расстояние между двумя точками соответственно с координатами (x_1, y_1) и (x_2, y_2). И это расстояние между точками-соседями должно быть меньше заданного малого числа. Запишем это условие:

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \le \varepsilon.$$

Основываясь на этом простом критерии, можно выстроить алгоритм отделения объекта, который назовем алгоритмом волны. Для описания алгоритма определим следующие понятия:

Множество объекта – множество точек, принадлежащих объекту.

Фронт волны – множество точек, принадлежащих объекту и обнаруженных на предыдущей итерации.

Итерация – анализ фронта волны на предмет обнаружения новых точек.

← – операция включения точки во множество.

Начинается анализ с *любой* непустой точки в пространстве. Так как эта точка непуста и мы ищем объекты любой природы, то она, очевидно, является точкой какого-либо объекта, хотя бы состоящего из нее одной. Для упрощения положим, что минимальное расстояние задано из каких-то внешних соображений.

```
Фронт волны ← Исходная точка
Множество объекта ← Исходная точка
Пока фронт волны не пуст, выполнять очередную итерацию
    Для каждой точки Фронта волны
        Для каждой точки из ε - окрестности точки Фронта
            Если точка из окрестности непуста
            и не принадлежит Фронту волны и множеству
            точек объекта,
                То Фронт волны ← точка окрестности
        Конец цикла
        Множество объекта ← точка Фронта волны
        Точку Фронта исключить из множества Фронта
    Конец цикла
Конец цикла итераций
```

Алгоритм волной проходит от *исходной* точки, присоединяя ко множеству объекта все непустые точки, до которых может дотянуться отрезком длины ε. Таким образом, в среде отделяется множество, которое можно признать объектом. Если в окружающем *пространстве* остаются точки, не включенные в данный объект, то это повод повторить процедуру для свободных точек.

Более тонкая процедура отделения

Способ отделения, рассмотренный нами выше, предполагает, что два объекта обязательно разделены *значительным* пространством. К сожалению, это довольно часто не так. Например: телевизор, стоящий на столе, цветы в вазе, люди, держащиеся за руки. Все это будет алгоритмом идентифицировано как один объект, поэтому нам необходимы более тонкие методы отделения.

Что человеку помогает отделить цветы от вазы или коробку, стоящую на столе, от стола? Две вещи: цвет и контурные линии. Пример первый – *два замкнутых контура* (рис. 5.5):

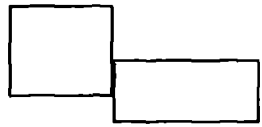

Рис. 5.5 ❖ Два замкнутых контура

Конечно, об этом рисунке можно сказать больше, чем просто «Здесь два объекта». Мы вполне можем утверждать, что здесь изображены два прямоугольника, один из которых сильно похож на квадрат. Однако задачу обозначения объекта мы пока не рассматриваем. А выделить именно два объекта можно на том основании, что мы видим два замкнутых контура, существование которых вполне можно установить уже известным нам, достаточно простым алгоритмом. Единственно, его придется усилить алгоритмом обнаружения общих участков контура, что является вопросом алгоритмической техники, а не принципа, но есть одно существенное возражение. Посмотрим на рис. 5.6.

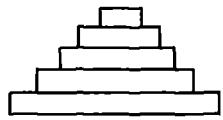

Рис. 5.6 ❖ Пирамида

Можно сказать, что нарисованы пять небольших прямоугольников, но можно утверждать, что нарисован единый объект, называемый пирамидой. На базе чего сделан более общий вывод? Очевидно, что такое утверждение (о пирамиде) стало возможным только после анализа, давшего наблюдателю общую закономерность – рисунок состоит из фигур одного типа, и эти фигуры плавно уменьшаются в размерах. Такой анализ возможен только на основе солидного объема знаний. Во-первых, необходимо уметь идентифицировать фигуры как одинаковые по типу, несмотря на разные размеры, во-вторых, наблюдатель должен уметь определить изменения в размерах как одинаковые или почти одинаковые. Сказанное означает, что умение увидеть в рисунке цельную фигуру никак не объясняется чистым наблюдением. Увидеть целое нельзя, можно лишь выделить целое после анализа, опирающегося на знания, а значит, этот пример к исследуемому вопросу – отделения объекта – отношения не

имеет. Кроме того, в любом случае сложный объект состоит из более простых, которые достаточно определять как области пространства, отделенные контуром.

Отделение цветом

Контурный рисунок в окружающем мире встречается достаточно редко, настолько редко, что можно сказать, что почти никогда. Любая реальная среда – это картина из пятен: разного размера, интенсивности, цвета. Для упрощения ситуации положим, что мир раскрашен оттенками серого, допущение полезное и мало что меняющее в задаче распознавания. Кстати, как утверждают биологи, кошки видят именно так, что не мешает им распознавать объекты окружающего мира с высокой эффективностью. А теперь посмотрим на рис. 5.7.

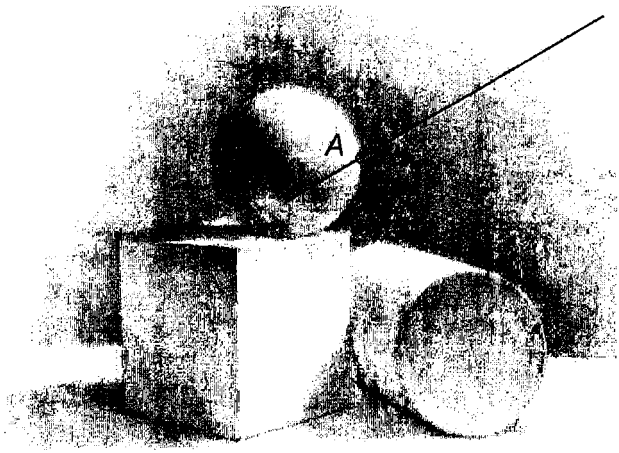

Рис. 5.7 ❖ Контрастный рисунок

Рисунок имеет ярко выраженные линии, отделяющие объекты, но они не проведены карандашом и линейкой. Здесь линии разделения появляются за счет резкого контраста. Для описания методов работы с изменениями цвета разработан достаточно мощный математический аппарат, но для понимания базовой идеи вполне достаточно обычных соображений здравого смысла.

Вернемся еще раз к рисунку, а точнее к черной линии, проведенной из правого верхнего угла. Проанализируем изменение цвета

вдоль обозначенной линии. Цвет фона вдоль линии плавно переходит из светло-серого в темно-серый. Но в точке *A* происходит резкий скачок оттенка из темно-серого обратно в светлый. Этот же скачок наблюдатель видит и в соседних точках, что создает эффект видимой линии, хотя никто эту линию циркулем не проводил. Линия отделяет окружность. Пройдем вдоль нее. Постепенно оттенок светло-серого переходит в темно-серый, но эффект линии цветового раздела остается. Внутри окружности насыщенность серого цвета меняется радикально, однако это не мешает воспринимать внутреннюю часть окружности как единый объект. Отсюда можно сделать вывод о том, что как в случае черно-белого, контрастного рисунка, так и в случае многоцветного наличие ограничивающих линий является фундаментальным фактором выделения объекта из окружающей среды.

Отделение областей пространства

В задаче анализа окружающего пространства можно выделить этап обнаружения макрообластей, то есть таких пространственных областей, которые, возможно, сами по себе не являются единым объектом (что будет выяснено наблюдателем только в ходе последующего анализа), но тем не менее сразу можно сказать, что нет объекта, принадлежащего двум и более областям. На рисунке, анализ которого только что проведен, три объекта находятся слишком близко, чтобы их можно было отделить, на рис. 5.8 ситуация иная.

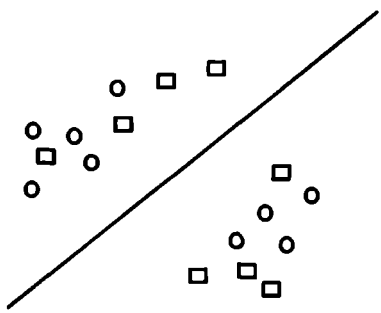

Рис. 5.8 ❖ Две области визуализации

Даже если бы не было разделяющей две группы объектов линии, хорошо видно, что здесь можно выделить две макрообласти, внутри

которых уже есть смысл поискать отдельные объекты. Рассмотрим несложный алгоритм выделения макрообластей. Первым шагом покроем пространство фигурами произвольной формы. Например, эллипсами (см. рис. 5.9):

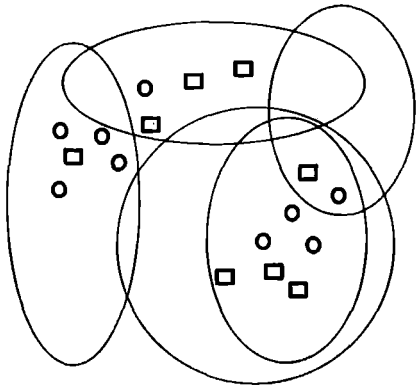

Рис. 5.9 ❖ Покрытие эллипсами

А теперь начнем отрезать от каждой области пустые части. Выполним эту процедуру в несколько шагов, чтобы более детально отследить процесс. Итак, шаг первый (рис. 5.10):

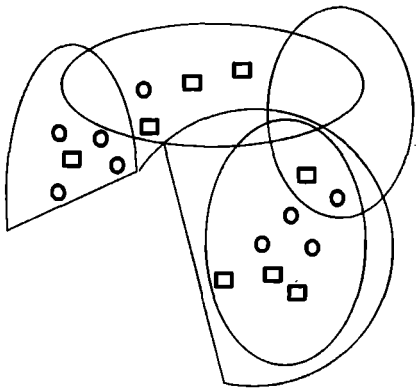

Рис. 5.10 ❖ Первый шаг отделения областей

Думаю, идея уже понятна. Продолжим процесс резки пустого рис. 5.11.

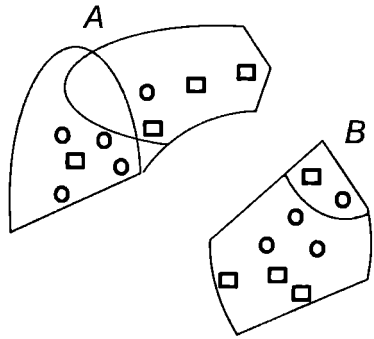

Рис. 5.11 ❖ Второй шаг отделения областей

После второго шага (или десятого, в зависимости от того, насколько детализирован процесс) остались две области, полностью изолированные друг от друга. Заметим, что область *A* несколько отличается от области *B*. В *A* есть две линии, отделяющие подобласти и не содержащие никаких объектов. Из этого следует, что внутренние линии в *A* можно просто убрать, а подобласти слить в одну. В макрообласти *B* ситуация немного иная. Есть две подобласти, и они обе содержат объекты. Заметим, что внутренние расстояния между объектами в подобластях и расстояния между объектами, лежащими в разных подобластях, практически не отличаются, следовательно, разумно эти подобласти тоже слить в одну. И получаем следующий результат на рис. 5.12:

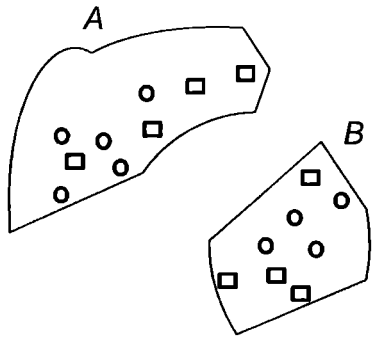

Рис. 5.12 ❖ Окончательный результат

Конечно, полученные макрообласти еще содержат значительные пустые пространства, а значит, их реальную форму еще можно уточ-

нить. Также понятно, что представленный алгоритм неформален и остается много важных вопросов. Например, необходимо продумать политику выбора линий, отсекающих пустые области, надо выбрать способ проведения линий, но это уже скорее вопрос техники, чем принципа.

Идентификация образа по шаблону

А далее займемся проблемой именования объекта. Пусть объект выделен, наш исследователь ориентирован на вполне определенный класс объектов, все остальное его не интересует, и если полученного объекта нет в списке известных, то тем хуже для него. В этой ситуации остается вопрос: как соотнести то, что видит наблюдатель, с известными ему названиями, как дать имя объекту? Рассмотрим простейший, но очень эффективный прием сравнения с шаблоном.

Шаблон – это некоторое идеальное, каноническое изображение. Оно включает в себя только существенные свойства объекта, можно сказать, что шаблон – это наиболее типичный представитель данного класса. Рассмотрим пример с буквой А.

Пять вариантов буквы А, имеющих право называться таковой буквой алфавита. Но все варианты достаточно различны, некоторые из них имеют какие-то элементы, черточки с претензией на украшение буквы. Наиболее разумно взять в качестве типичного представителя изображения буквы четвертый вариант. В этой букве «А» нет ничего лишнего, вариант минималистический.

Каким образом можно использовать шаблон для идентификации выделенного объекта? Очевидно, совмещением с идентифицируемым объектом. Однако трудно предположить, что шаблон хорошо и точно ляжет на образ. Всегда наложение окажется неточным, не вполне совпадающим. А значит, встает задача оценки погрешности. Рисунок 5.13 иллюстрирует самую простую ситуацию, в которой оценить погрешность труда не составляет.

На рисунке две буквы написаны разными шрифтами и затем совмещены. Если теперь на совмещенном рисунке выбросить черные области, не совпадающие с красными, и красные области, не совпадающие с черными, то оставшееся изображение, очевидно, будет буквой А, а значит, разница между этими двумя изображениями находится

в пределах допустимой погрешности. К сожалению, такой прямолинейный метод наложения хорош только в самых простых случаях. Посмотрите на рис. 5.14.

Рис. 5.13 ❖ Совмещение шаблона

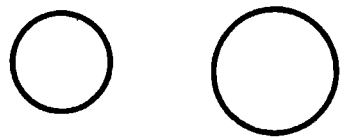

Рис. 5.14 ❖ Несовмещаемые окружности

Мы, конечно, понимаем, что это две окружности. Но предположим, одна из них является шаблоном, а вторая – идентифицируемым изображением. И как бы мы не накладывали одну на другую, всегда погрешность (количество несовпадающих точек) будет слишком велика. Но тем не менее это две окружности. Что делать в таком случае? Мы понимаем, что окружностей разного размера сколь угодно много и создавать шаблоны под каждую невозможно.

Метод малых преобразований

Пусть есть один шаблон окружности и к нему привязано преобразование, изменяющее размер образа окружности во столько раз, во сколько раз радиус образа больше или меньше радиуса шаблона. Тогда вполне возможно обойтись только одним шаблоном, действуя следующим образом.

1. Мы получаем для анализа образ некоего объекта.
2. Мы полагаем, что это окружность. Если объект не окружность, то после неудачи сравнения шаблона и образа мы перейдем к шаблону другого класса объектов.
3. Следующим шагом необходимо определить радиус образа. По определению, радиус – это расстояние от центра до любой

точки окружности, следовательно, задача расчета радиуса сводится к задаче определения центра окружности. Тогда, до начала работы этого алгоритма, мы должны были определить область объекта. Выделим в этой области максимальный квадрат. Центр окружности – это центр квадрата, а ее радиус – половина стороны квадрата.

4. Изменим радиус шаблона до вычисленного и приложим шаблон к образу.
5. Если хорошего совпадения нет, то это еще не означает неудачи. Возможно, есть небольшая ошибка в определении центра. Выполним некоторое количество попыток сдвига центра шаблона в разных направлениях на небольшую величину.
6. Если сдвиг центра не дает нужной точности, то изменим на малую величину радиус шаблона и повторим предыдущий пункт.
7. Пункты 5 и 6 нет смысла повторять до бесконечности. Можно ввести некую эвристическую договоренность, что, скажем, 10 попыток вполне достаточно. Все равно работа такого рода алгоритмов имеет вероятностную, эвристическую природу.

А теперь рис. 5.15.

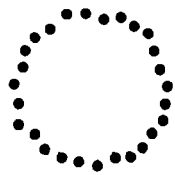

Рис. 5.15 ❖ Плохая окружность

Наш алгоритм не сработает на этом примере. Оставим от него только один пункт – определение радиуса и центра окружности. А вот дальнейшие действия строятся на иной идее. Если исследуемый образ есть окружность, то на расстоянии радиуса от центра должна быть точка. Но так как радиус определен приблизительно и центр окружности тоже определен приблизительно, то точка окружности с высокой долей вероятности находится где-то в малой окрестности от ожидаемой точки. Определение радиуса малой окрестности – опять-таки эвристическая процедура, но можно утверждать, что радиус малой окрестности должен быть существенно мал, в сравнении с радиусом исследуемого объекта. Идея, думаю, ясна, записывать детальный алгоритм не будем.

Идея преобразования в общем виде

На примере окружности мы рассмотрели только один вид преобразования – подобие. Это наиболее простое преобразование, его необходимость совершенно очевидна. Можно привести массу примеров, в которых разные по размеру объекты принадлежат к одному и тому же классу. Поэтому процедура нормализации (так называется процедура приведения объекта к стандартному размеру) очевидна, но также очевидно, что размер – не единственное препятствие для распознавания образов. Рисунок 5.16 хорошо иллюстрирует это утверждение.

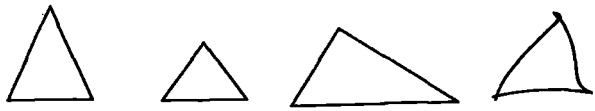

Рис. 5.16 ❖ Неподобные треугольники

На рисунке четыре треугольника. Допустим, первый из них представляет собой шаблон. Ясно, что этот шаблон не совпадет ни с одним из оставшихся, и точно так же понятно, что проблему не исправить только преобразованием подобия. Однако геометрия может дать нам целый набор преобразований, среди которых можно найти и такие, которые смогут совместить оставшиеся три с шаблоном. Не будем детально рассматривать, как это возможно, каждый из нас может придумать такие преобразования без строгой математической формулировки. Этот пример наводит на следующую идею: задачу распознавания методом сравнения с шаблоном можно сформулировать как задачу поиска допустимого преобразования объекта (или шаблона, кому как нравится), такого, что наложение дает погрешность в допустимых пределах.

Здесь важно уточнение насчет допустимых преобразований. Современная геометрия дает настолько большой набор преобразований, что на самом деле можно любой геометрический объект преобразовать в любой, что делает задачу распознавания бессмысленной. Например, топология (наука, являющаяся некоторым обобщением геометрии), разрешающая все преобразования, кроме склеивания и разрезания, не отличает прямоугольник от окружности. Существует топологически законное преобразование, преобразующее прямоугольник в окружность и наоборот! Поэтому уточним, что допустимое

преобразование – это преобразование, известное обычному человеку, не геометру. Плюс к этому можно и нужно добавить, что преобразование не должно изменять объект слишком сильно. А что такое «слишком сильно» и «не слишком сильно», мы опять вынуждены решать на основе некоего эвристического допущения.

Распознавание объекта по набору признаков

В отношении предыдущей задачи распознавания окружности заметим одну очень важную вещь: если каким-то образом установлено, что существует точка, такая, что все точки объекта равноудалены от нее, то никакого шаблона на самом деле накладывать уже нет необходимости, уже и так понятно, что перед нами окружность. Это означает, что знания об объекте могут храниться в памяти исследователя не в виде шаблона, а в виде набора признаков, однозначно (ну, или с высокой степенью вероятности) определяющих объект. Различные науки, и не только геометрия, дают нам признаки визуальных объектов. Наверное, в еще большей степени таковую информацию люди получают из своего опыта наблюдения. Человеческий опыт вполне можно передать машине, что даст довольно много. Конечно, нужен специальный язык для записи формализованного знания о признаках, но это скорее технический момент. А есть в этом деле момент принципиального характера.

Классы объектов могут пересекаться, вкладываться друг в друга. Например, класс яблоки и класс груши входят в класс фрукты. И надо полагать, что на том основании, что два класса входят в более широкий, у них должны быть общие признаки, присущие общему классу. Из этого соображения следует, что признаки, как и определяемые ими классы, можно разбить по уровням абстракции, что, в свою очередь, дает возможность оптимизировать процесс распознавания. Процесс распознавания квадрата мог бы составить следующие шаги.

Шаг 1. Определим, представляет ли данный для наблюдения образ часть плоскости, ограниченную линией. Линия при этом может представлять собой действительно линию проведенную, ну, например, карандашом, а может быть, и просто линией разделения двух цветов или оттенков.

Шаг 2. Если образ есть область из предыдущего шага, то посмотрим, состоит ли линия, ее ограничивающая, из прямых отрезков. Если

да, то можно прийти к заключению, что перед нами многоугольник. Еще нет, не квадрат, еще даже не правильный многоугольник и даже пока не выпуклый, но уже многоугольник, а значит, есть смысл продолжать пытаться распознать квадрат.

Шаг 3. Если таковых отрезков ровно четыре, то исследуемый образ – безусловно, четырехугольник. В этом случае:

Оставшиеся шаги. Выясним, являются ли отрезки предыдущего пункта попарно параллельными и одинаковой ли они длины. Если оба условия выполняются, то, очевидно, исследуемый объект – квадрат.

Проведенный анализ – пример целенаправленного поиска квадрата среди наблюдаемых объектов. Если нет возможности поиска целевого образа, входящего в четко заданный класс, то схема действий может быть такова:

Пункт 1. Выделяем из полученного образа очередной признак. Не важно, какой, может быть, то, что первым бросается в глаза.

Пункт 2. Находим абстрактный класс, заданный выделенным признаком. На этом шаге процесс можно прекратить и в качестве решения взять любой объект из полученного класса.

Пункт 3. Если же вариантов слишком много (объем класса очень велик) и такая степень детальности наблюдателя не устраивает, то выделяется следующий признак, уточняющий и конкретизирующий класс объектов. Анализируя рисунок, состоящий из множества линий, выделив несколько линий и выяснив, что они состоят из прямых отрезков, мы тем самым отсекаем возможность распознавания криволинейных фигур, а стало быть, отсекаем возможность прийти к образу эллипса, но получаем гипотезу, что перед нами, возможно, многоугольник.

Метод, сравнивающий шаблоны, более эффективен, если речь идет о простых объектах, так как процедура наложения шаблона сама по себе достаточно проста, при условии, конечно, что грамотно определен набор допустимых преобразований. Если же наблюдаемое изображение сложно, содержит множество визуальных элементов, то метод выделения признаков более предпочтителен, хотя он может оказаться причиной так называемого обмана зрения. Есть тому хорошие примеры. Их суть в том, что в одной и той же области наблюдения можно обнаружить признаки совершенно различных и совершенно не сопоставимых друг с другом объектов. Вопрос в том, под каким углом вы смотрите и какие элементы обнаруживаете первыми. Вот только два классических примера (рис. 5.17):

Рис. 5.17 ❖ Два изображения на одном рисунке

Слева можно увидеть белую вазу, а можно два черных человеческих лица, глядящих друг на друга, а справа можно увидеть старуху, а можно молодую девушку. Совершенно не понятно, каким образом хотя бы один из образов можно было бы обнаружить с помощью шаблонов, а анализ признаков – единственно доступный метод в таких относительно сложных ситуациях – дает многозначный результат.

Еще одна существенная проблема состоит в том, что между анализирующим устройством и внешней, наблюдаемой средой стоят органы восприятия информации, в нашем случае глаза, и устройство предварительной подготовки изображения, его интерпретации в том формате, который понятен анализатору. И органы восприятия и тем более устройство, интерпретирующее изображение, информацию искажают. В результате анализ может дать ответ, не соответствующий действительности. Ниже еще два классических примера (рис. 5.18):

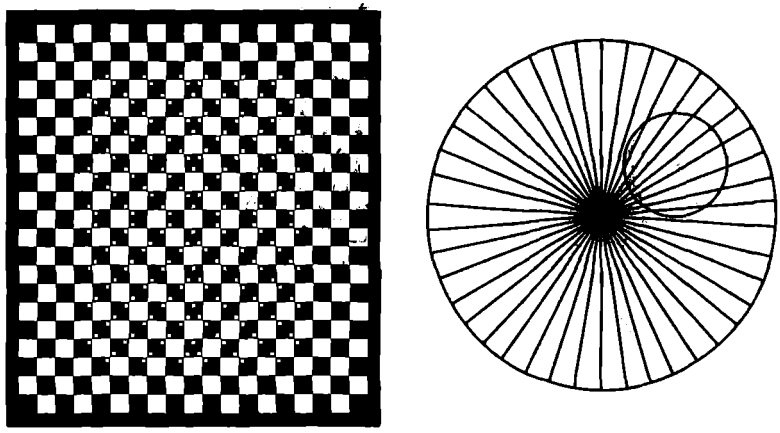

Рис. 5.18 ❖ Обман зрения

На рисунке слева совершенно четко видно, что прямоугольники криволинейные, однако все линии рисунка совершенно прямые. На рисунке справа идеальная окружность выглядит сильно искаженной. Конечно, можно возразить, что это проблемы человеческого восприятия. Да, это так, но человеческий механизм восприятия позволяет решать очень сложные задачи распознавания объектов, и, следовательно, его нельзя признать несовершенным. И подумайте вот о чем. А возможен ли идеальный, не искажающий информацию воспринимающий аппарат? Лично я полагаю, что нет. Передача изображения на анализ без искажения потребует всего огромного массива никак не упорядоченной информации. Такой массив данных перегрузит анализатор настолько, что он не сможет дать не только правдоподобного описания увиденного, но не сможет дать вообще никакого описания за реальное время. А любая интеллектуальная система, действующая в реальной среде, обязана вырабатывать не только адекватную реакцию, но и вырабатывать ее за вполне определенное, зачастую очень ограниченное время.

Отсюда следует главный вывод. Любой метод распознавания, как и любой метод, касающийся интеллектуальных задач, вынужденно носит эвристический характер, что, впрочем, для нас неудивительно, эта мысль красной нитью проходит через все главы читаемой вами книги.

В заключение

В начале главы было упомянуто, что в распознавании нуждаются не только визуальные объекты. Дайте немного поговорим об этом. Предположим, на анализ поступил следующий текст:

О сколько нам открытий чудных
Готовит просвещенья дух
И опыт сын ошибок трудных
И гений парадоксов друг

Здесь возникает целый ряд вопросов: откуда известно, что перед нами литературный текст, откуда известно, что это литературный текст в стихотворной форме, и как определить, что это стихи, принадлежащие перу Александра Сергеевича Пушкина? Последний вопрос решается просто, а вот как ответить на два первых вопроса? Или еще один пример:

Представим себе совокупность из N+1 урн, в каждой из которых содержится N шаров; урна с номером k содержит k красных и N – k черных шаров. Случайным образом выбирается урна, и из нее n раз извлекаются шары...

Ясно, что этот текст уже не литературный. Это уже область математики, а точнее, теории вероятности, и, возможно, текст представляет собой часть условия некоей задачи. И опять тот же вопрос: каким образом можно прийти именно к такому заключению. Поиск ответа на такого рода вопросы тоже можно интерпретировать как задачу распознавания. И эта задача неизмеримо сложнее распознавания визуальных объектов.

Начнем с того, что смысл имеет целое множество представлений. Самая простая и распространенная форма – это речь. Речь может быть устной и письменной. Смысл может передаваться визуально в виде графического изображения. Каковым, например, является картина, написанная художником. Нарисованная картина, очевидно, несет в себе не только набор визуальных объектов, в ней важны взаимодействия между ними, культурный контекст, без которого картину нельзя понять. В смысле задачи визуального распознавания картина Малевича, которую так и называют – квадрат Малевича, не представляет собой ничего, кроме черного квадрата. Но такое понимание совершенно не объясняет ее огромной популярности. Мона Лиза Леонардо да Винчи системой распознавания будет определена как женский портрет, что также не объясняет ее огромной культурной ценности. Видимо, дело здесь именно в культурном контексте.

Текстовая форма смысла может быть прозой, может иметь стихотворную форму, смысл может переноситься в форме музыки. Все это многообразие форм является предметом задачи распознавания. Но и это еще не все. Смыслы разбиваются на области знания, смыслы связаны между собой исторически, логически и т. д. Распознание смысла обязательно включает в себя определение его места в системе знания.

Задача кажется необъятной, но к ней есть ключик. Этот ключик заключен в термине «понятие». Наше мышление в каком-то смысле можно назвать процессом оперирования понятиями разной степени абстракции. Есть понятийный аппарат наиболее общий, так сказать, общемыслительный, есть понятийный аппарат специфический, используемый в конкретных областях знания. Отсюда идея – задача распознания смысла сводится к выделению в информационном со-

общении терминов, обозначающих понятия. Выделив такой термин, исследователь сразу получает столько информации о смысле, сколько можно извлечь из его базы знаний, каковая как раз и представляет собой систему связанных между собой различными отношениями понятий.

Этой идеей я позволю себе закончить тему, так как распознавание смыслов – проблема, и по методике, и по идеологии настолько сильно отличающаяся от распознавания визуальных образов, что здесь более уместна отдельная глава, кроме того, задача распознавания смыслов в значительной мере эквивалентна общей задаче построения системы искусственного интеллекта, то есть эквивалентна задаче, которой посвящена вся эта книга. Еще в заключение позволю себе напомнить, что задача распознавания образов рассматривается в главе, посвященной нейронным сетям. Отличие этой главы в том, что нейронные сети не есть чистая теория, а, скорее, техническая реализация, здесь же мы попытались рассмотреть некоторые идеи, никак не привязанные к каким-либо техническим схемам и устройствам.

Глава 6

Искусственное познание

Большинство специалистов по искусственному интеллекту сходится во мнении на эту область знания, полагая ее отраслью инженерии, ставящей своей целью моделирование такого поведения, которое в человеческом исполнении считается разумным. Многие крупные ученые высказываются даже в таком ключе, что иное представление об ИИ не имеет смысла. Кроме того, уже сейчас инженерный подход дает настолько серьезные выгоды от внедрения технологий ИИ, что в правильности общей линии сложно сомневаться. Однако ясно, что сколь бы далеко мы не продвинулись в создании систем, виртуозно и даже лучше человека решающих отдельные задачи, эти системы можно будет назвать интеллектуальными только с некоторой натяжкой, не забывая приставки «искусственный». Есть несколько проблем, решение которых пока достаточно туманно.

Камни преткновения на пути искусственного интеллекта

Мы не только умеем мыслить,
мы умеем ставить цель
для своего интеллектуального аппарата

Дело в том, кто и как ставит цель. Человек в этом вопросе достаточно свободен. И если даже отдельному индивидууму, или социальной группе, или даже целому народу (что в истории уже было, и не раз) навязываются какие-то цели, то такие цели носят характер глобаль-

ных, мировоззренческих. Ни одна идеология не предписывает человеку, что ему делать в той или иной конкретной ситуации, как строить свой день, выполнять трудовые или бытовые операции. Существует много целей, в формулировке которых каждый человек достаточно свободен и зачастую нелогичен и даже неразумен с чьей-то точки зрения. В принципе, под воздействием каких-то внутренних мыслительных процессов любой из нас может отказаться и от целей навязанных, если они не поддерживаются обыкновенным физическим насилием. В общем, можно с уверенностью утверждать, что мы в вопросе целеполагания свободны.

С искусственным интеллектом дело обстоит иначе. ИИ, способный доказывать геометрические теоремы, не сможет играть в шахматы, а ИИ, управляющий системами ПВО, не сможет поставить медицинский диагноз. Конечно, можно возразить, что даже очень хороший геометр, весьма вероятно, не умеет играть в шахматы, а самый лучший врач не сможет управлять простым зенитным орудием, но дело в том, что врача можно обучить другой профессии, возможно, он и не станет в иной ипостаси классным профессионалом, но обучить можно. Кроме того, врач, учитель, музыкант, строитель могут захотеть научиться чему-то иному, чего не скажешь о самой лучшей программе. И из этого утверждения мы выходим на следующую проблему:

Человек не просто обучаем, он универсально обучаем, каждый из нас может достичь чего-либо в любой области и довольно многого в разных. Человек, обучившийся одной профессии, в течение жизни может поменять ее на другую

Каждого психически полноценного человека можно обучить любой области знания. Этот факт демонстрирует средняя школа, включающая в свою программу все основные познавательные предметы. Конечно, есть такие понятия, как способности и талант, но эти вещи создают успешность в освоении, а не возможность. Для современной техники эта человеческая способность пока за пределами достижимого. Переобучаемость человека в программистских терминах означает возможность для системы ИИ перепрограммировать саму себя. И здесь виден целый ряд проблем.

Это и вопрос инструментария, используемого для создания программных систем. В нашем распоряжении пока только языки программирования, которые вряд ли подойдут для создания программы, способной писать программы. Это ясно из следующего соображения:

все современные алгоритмические языки, как императивные, так и декларативные, можно рассматривать как методы преобразования последовательностей символов. Искомая же система должна уметь оперировать знаниями, что бы этот термин не означал. Можно предположить, что такая система должна обладать языком, сопоставимым по своим выразительным возможностям с естественным. А язык – не самое большое затруднение. Человек владеет очень серьезными мыслительными инструментами, например может мыслить индуктивно, обобщая частные события в общие правила.

Человек умеет делать выводы, переходя от общего к частному, но, что еще более интересно, мы умеем выполнять и обратный переход от частного к общему, и этот переход совсем не выглядит хорошо алгоритмизируемым действием

Техника так называемого правильного мышления, гарантирующего качественный результат, была целью человеческой науки и философии с того самого момента, как наука и философия оформились в общественном сознании человечества. И что любопытно, быть может, самый существенный прорыв в разработке мыслительных методов был совершен еще в древнем мире с созданием логики силлогизмов. В первой главе этой книги есть краткий набросок системы Аристотеля. Дедуктивный метод Аристотеля основан на очевидном положении. Если есть некоторое утверждение, имеющее характер всеобщности, то оно истинно для всех частных случаев. Если мы имеем большой класс объектов, например яблоки, и утверждение обо всем классе, например «Яблоки съедобны», то это утверждение является всеобщим. Но, согласившись с общим утверждением, мы вынуждены соглашаться и с частным утверждением относительно каждого конкретного яблока. Такой ход рассуждений называется дедуктивным, его алгоритмизируемость очевидна.

В получении истинных суждений можно пойти и несколько другим путем. Будем находить отдельные экземпляры яблок и экспериментально устанавливать их съедобность. Вопрос вот в чем. Можно ли, проведя некоторую серию экспериментов, с полной определенностью заявить об истинности общего утверждения «Все яблоки съедобны»? Сколько для этого необходимо экспериментов? Если установлена съедобность 100 яблок, можно ли утверждать то же самое относительно 101 яблока? Да, но лишь с некоторой степенью вероятности. Если проведено 1000 экспериментов, то правдоподобность этого утверждения относительно 1001 яблока увеличивается, но не до абсолют-

ной истины. В общем, вне зависимости от количества экспериментов обобщающее (индуктивное) утверждение будет лишь правдоподобным в той или иной степени, но не безусловно истинным. И это создает проблему.

Безусловный дедуктивный переход легко формализовать и втиснуть в искусственную систему принятия решений. С индукцией, хотя она и выглядит всего лишь операцией, обратной дедукции, так не получается. Индуктивные выводы всегда содержат в себе некоторую степень неопределенности, и процедура принятия решения на основании индуктивного вывода далеко не так очевидна.

А мы еще умеем мыслить объект, не наблюдаемый непосредственно. Это то, что называется абстрактным мышлением, – форма мышления о возможном, форма мышления о всеобщем, но не о том, что перед глазами

Еще одна интересная возможность человеческого мышления – это умение отделять свойства объекта от носителя, выделяя только наиболее существенные. Мы можем рассуждать о красном или теплом, не имея в виду красную розу или теплые руки. Высшая форма абстракции – это «понятие». Понятие является представлением в языке большой группы объектов, не привязанным к конкретному представителю группы. При этом объекты группы могут иметь совершенно разную природу. Понятие «круглое» можно применить к объектам: стол, мяч, планета, сумма денег и т. д.

Есть, кроме того, в нашем сознании и такие понятия, за которыми не стоит никакого материального объекта вообще. Например: привязанность, чувство долга, удовольствие. Человеческое мышление выработало очень сложную, иерархическую систему понятий, и любая система ИИ, претендующая на звание умеющей мыслить, должна быть способна оперировать абстракциями. И конечно, системы логического вывода, различные эвристические методы поиска решения, нацеленные на борьбу с полным перебором, не имеют ничего общего с настоящим абстрактным мышлением.

И наконец, человек умеет мыслить себя, создавая представления о том, что он есть сам и как его существование соотносится с существованием окружающего мира

Процесс познания – самая главная функция человеческого мышления – глубоко субъективен. Хорошо это или плохо, но это так. А субъ-

ективное мышление направляется не только новой информацией, но и уже накопленным знанием, и, даже более того, своим отношением к тому, что уже известно. Побочным, но принципиальным следствием этого факта является то, что познающий субъект одной из целей познания обязательно ставит определение своего места в этом мире и определение самого себя. В каком-то смысле это явление рефлексии есть следствие свободы мышления. Если мышление вольно выбирать себе объект для исследования, то оно имеет право выбрать для изучения и самое себя. А значит, способность к самоанализу должно признать важным свойством интеллектуальной системы.

Полагаю даже, что это свойство самое фундаментальное и самое важное. В спорах об искусственном интеллекте всегда есть вопрос – что значит для ИИ быть подобным человеку? Мы уже видим, что способность решать сложные задачи не является гарантией подобия, и способность обучаться, например, шахматам, не приближает машину к человеку. Ведь даже примитивные нейронные сети, создаваемые человеком, способны обучаться решению каких-то классов задач. Может, разгадка принципиального различия между человеком и возможной машиной с ИИ заключается в способности человека осознавать себя? А что, если это действительно так?! Правда, надо признать, что в этом случае так же, как и вообще в вопросе интеллекта, встает проблема определения, и вряд ли определить термин «сознание» будет проще, нежели термин «интеллект».

Единственная твердая опора для построения систем искусственного интеллекта – формальная логика – слишком ограничена и не способна объяснить возможности разума человека

Первые идеи создания ИИ упирались в моделирование логического вывода. Однако довольно быстро прояснилась ограниченность такого подхода. Строгий логический вывод как метод получения нового знания начинает хромать даже в задачах доказательства теорем. Такой вид интеллектуальной деятельности, на первый взгляд, кажется весьма привлекательным, и в хорошо формализованных системах его реализация достаточно проста. Действительно, если есть набор исходных аксиом и способ вывода теорем, дающий гарантированный результат, тогда доказательство теоремы сводится к следующему алгоритму:

Шаг 1. Формулируем множество истинных утверждений. На первом шаге оно совпадает со множеством аксиом теории.

Шаг 2. Выводим из множества истинных утверждений все возможные теоремы.

Шаг 3. Добавляем все выведенные теоремы во множество истинных утверждений и повторяем второй шаг.

В этом алгоритме для его завершения достаточно предусмотреть еще один пунктик – проверку, нет ли в уже построенном множестве истинных утверждений, искомой теоремы, и если да, то процесс вывода можно прекратить. А теперь предположим простейшую ситуацию: множество исходных аксиом состоит только из двух утверждений. И пусть из пары истинных утверждений можно получить только одну теорему. Это означает, что на первом шаге множество истин будет состоять из трех утверждений (две аксиомы и одна теорема), на втором – уже из пяти (добавятся две теоремы), и далее рост пойдет очень быстро. Конечно, механическое соединение двух истин не всегда даст теорему, но с ростом теорем количество продуктивных пар также будет расти. Реально даже в небольшой системе аксиом, которая может состоять из значительно большего числа утверждений, рост числа теорем будет идти так быстро, что множество истинных утверждений очень быстро станет необозримым, быть может, так и не дойдя до требуемой теоремы.

Для разрешения этой, не самой сложной проблемы уже в «Логик-теоретик» – первой программе, умеющей доказывать теоремы, ее авторы ушли от построения полного дерева перебора, заменив его правдоподобными эвристиками. А есть и более сложные вопросы. Например, что делать, если неизвестно, существует ли вообще доказательство теоремы, ведь если нет, то даже укороченное эвристиками дерево вывода будет расти неограниченно.

Намного более сложные вопросы возникают, если мы уходим от таких простых теорий, как исчисление высказываний, в рамках которой работал «Логик-теоретик». А в человеческой науке есть области знания, которые просто нельзя выстроить как набор следствий из ограниченного набора аксиом. Такими, например, являются все естественные и гуманитарные науки, такими, по сути, являются все человеческие науки, за исключением некоторых отраслей математики.

В общем и целом необходимо признать, что в рамках формальной логики мало что можно. Система, претендующая на звание разумной, должна уметь пользоваться внелогическими инструментами для принятия решения о направлении исследований. Необходимо признать, что, приступая к поиску решения задачи, мы имеем огромное мно-

жество вариантов, среди которых правильные встречаются крайне редко, и метода, позволяющего идти точно к требуемому решению, минуя все ошибочные, просто не существует. Любое направление поиска в таком множестве носит характер более-менее оправданного, сводясь в предельном случае к полному перебору.

Логический вывод и доказательство теорем

О машине, решающей задачи, мечтал еще Лейбниц, но, конечно, механическая элементная база, возможная в то время, могла дать, максимум, устройство, подобное арифмометру, и не более того. Но уже первые ЭВМ архитектуры фон Неймана изменили ситуацию кардинально. Оказалось, что даже их весьма ограниченных возможностей было вполне достаточно для моделирования логического вывода, хотя бы и в исчислении высказываний. Такая программа была создана А. Ньюэллом, Дж. Шоу, Г. Саймоном еще в 1956 году. Пожалуй, это была первая попытка проникнуть в тайны мышления, и уже этими авторами была показана насущная необходимость разработки эвристических механизмов. Программа «Логик-теоретик», которую еще иногда называют машиной доказательства теорем, в разработке теорем исчисления высказываний по своим методам выходила далеко за пределы этой теории. Сейчас мы попытаемся понять сильные стороны идеи этой тройки ученых, но для тех, кто, быть может, не знает, что такое исчисление высказываний, необходим краткий экскурс в теорию.

Высказывание – это краткое осмысленное предложение, про которое можно сказать, что оно истинно или ложно. Истинность или ложность является обязательным свойством высказывания, даже если именно сейчас нет возможности установить, что имеет место быть. Например, «В Африке есть не менее десяти горных вершин высотой более 4 тысяч метров». Немного найдется людей, которые сразу могут сказать, так это или не так, но мы понимаем, что, вооружившись соответствующими справочниками и потратив некоторое время, установить, так это или нет, можно. А то, что высказывание «Все люди планеты Земля имеют рост более двух метров» ложно, ясно сразу, без дополнительного исследования. Точно так же ясно, что высказывание «Хлеб – это пища» истинно. Истинность высказывания «На Марсе есть жизнь» проверяемо, в принципе, но пока мы не имеем техноло-

гической возможности это установить, а предложение «Лев Толстой как писатель лучше Джека Лондона» не проверяемо, так как нет критерия, позволяющего сравнивать двух мастеров литературы, а значит, это не высказывание. В общем, мы должны быть уверены, что существует способ проверить истинность кандидата на высказывание, и не важно, можем ли мы выполнить проверку уже сейчас. Возможность достаточна.

Исчисление высказываний называется математическим термином «исчисление», по той причине, что логикам удалось создать теорию, способную с помощью небольшого количества правил и законов определять истинность высказывания и получать новые истинные высказывания из системы уже имеющихся

Приведенные выше примеры называются элементарными высказываниями. Из них с помощью логических операций можно строить сложные высказывания. Логические операции, применяемые к элементарным высказываниям, создают новые высказывания. Пример такой операции – отрицание, «Хлеб – это не пища» – отрицание элементарного высказывания «Хлеб – это пища». Логических операций можно придумать много. Но для дальнейшего рассказа нам хватит только нескольких из них. Пример отрицания уже приведен. Отрицание – это так называемая одноместная операция, то есть применяемая к одному высказыванию, все остальные операции двуместные, то есть соединяющие два высказывания в одно сложное.

Логическое сложение, оно же дизъюнкция, оно же операция ИЛИ (общепринятое обозначение $\vee$*).* Соединяет два высказывания в следующей форме: *А ИЛИ В*, где *А*, *В* – высказывания. Дизъюнкция истинна, если истинно хотя бы одно из входящих в нее высказываний, и ложна, если ложны оба. «Кошка умеет говорить» ИЛИ «Человек умеет говорить» – второе высказывание истинно, а значит, истинно и сложное высказывание.

Логическое умножение, оно же конъюнкция, оно же операция И (общепринятое обозначение $\wedge$). Соединяет два высказывания в следующей форме: *А И В*, где *А*, *В* – высказывания. Конъюнкция истинна, если истинны оба высказывания, и ложна, если ложно хотя бы одно. «Кошка умеет говорить» И «Человек умеет говорить» – первое высказывание ложно, а значит, ложно и сложное высказывание.

Логическое следование, оно же импликация, оно же операция $\rightarrow$. Соединяет два высказывания в следующей форме: $A \rightarrow B$, где *А* (по-

сылка), *B* (следствие) – высказывания. Здесь вопрос истинности решается сложнее. Если посылка истинна и следствие истинно, то импликация истинна. Если посылка ложна, то импликация истинна независимо от истинности следствия. Такое определение импликации выражает тот факт, что из истины должна следовать истина, а из лжи может следовать все, что угодно. Импликация ложна, если утверждает, что из истины следует ложь. Примеры: «Этот предмет – яблоко», следовательно, «этот предмет съедобен». «Кошка не умеет говорить», следовательно, «Человек умеет говорить». Проанализируем утверждения. В обоих высказываниях первой импликации есть указание «Этот», что означает наличие некоего предмета и его доступность для проверки. Значит, оба предложения можно признать высказываниями с проверяемой истинностью. В этом случае истинность импликации определяется истинностью входящих в него простых.

Вторая импликация, очевидно, истинна (истинны оба входящих в нее высказывания), но она соединяет два не связанных между собой факта. Из неумения кошки говорить никак не следует умение человека, но мы это знаем не из теорем исчисления высказываний, а из своего внелогического знания. А значит, вторую конъюнкцию также следует признать законной, ее осмысленность – за пределами нашей теории.

Можно еще привести примеры логических операций, но для дальнейшего изложения этих достаточно, заметим только, что любое высказывание, в том числе и сложное, полученное объединением элементарных логических операций, имеет лишь два значения: истину и ложь. Истину принято обозначать единицей, ложь – нулем, а связь между значениями элементарных высказываний и значением сложного выражают таблицами истинности. Ниже приведена общая таблица для трех двуместных операций:

A	*B*	$A \vee B$	$A \wedge B$	$A \rightarrow B$
Истина	Истина	Истина	Истина	Истина
Истина	Ложь	Истина	Ложь	Ложь
Ложь	Ложь	Ложь	Ложь	Истина
Ложь	Истина	Истина	Ложь	Истина

Пользуясь этой таблицей, легко составить таблицу истинности для любого сложного высказывания. Приведем пример: $(A \vee B) \rightarrow B$.

A	B	$A \vee B$	$(A \vee B) \to B$
Истина	Истина	Истина	Истина
Истина	Ложь	Истина	Ложь
Ложь	Ложь	Ложь	Истина
Ложь	Истина	Истина	Истина

Существуют сложные высказывания, значения которых при любой комбинации значений элементарных являются истинными. Проиллюстрируем это таблицей истинности на следующем примере: $(A \wedge B) \to B$.

A	B	$A \wedge B$	$(A \wedge B) \to B$
Истина	Истина	Истина	Истина
Истина	Ложь	Ложь	Истина
Ложь	Ложь	Ложь	Истина
Ложь	Истина	Ложь	Истина

Такое высказывание можно назвать аксиомой, так как именно аксиомы обладают свойством тождественной истинности. Конечно, в исчислении высказываний существует не одна аксиома, а целая система. Если мы сформулируем несколько аксиом и добавим к ним средства логического вывода, позволяющие из имеющихся аксиом и уже доказанных теорем получать новые теоремы (сложные истинные высказывания), то это будет означать, что мы получим возможность развивать некоторую логическую теорию.

Конечно, теорию, полученную таким путем, нельзя назвать содержательной, так как сами высказывания не несут в себе никакого интересного смысла. И столь примитивная форма, соединяющая простейшие высказывания простейшими логическими операциями, также не дает надежды на получение сложных теорий, однако сама возможность формализации логического вывода весьма интересна. Но, прежде чем заняться вопросами формализации, нужен еще метод логического вывода. Самая общая схема выглядит так:

$$\frac{A \qquad A \to B}{B}.$$

Читается схема так: если посылка умозаключения истинна (высказывание A) и истинна импликация $(A \to B)$, то, очевидно, истинно

и заключение (высказывание B). Мы уже упоминали, что импликация может быть истинна при ложной посылке, но такая ситуация не интересна, а из истины должна следовать истина. Поэтому если посылка истинна и есть возможность установить истинность импликации, то и заключение будет обязательно истинным. Этот простой общий принцип является основой для нескольких приемов получения истинных высказываний.

Набор правил вывода машины «Логик-теоретик»

Авторы программы, которую мы в дальнейшем будем называть машиной, использовали три правила вывода:

- *правило подстановки*. Любую переменную (элементарные высказывания в дальнейшем будем называть переменными) в выражении можно заменить на любое выражение, при условии что эта замена выполняется везде, где встречается заменяемая переменная. Очевидно, что если исходное выражение представляет собой истинную теорему (тождественно истинное высказывание, или, иначе говоря, выражение истинно при любых значениях входящих в него переменных), то после замены мы также получим истинную теорему;
- *правило замены*. Если таблица истинности выражения A, входящего как составляющее в выражение B, совпадает с таблицей истинности некоторого другого выражения C, то, очевидно, A можно заменить на C, не меняя истинности исходного выражения B;
- *правило отделения*. Если выражения A и $A \rightarrow B$ являются истинными высказываниями, то B также истинно. Это правило – не что иное, как схема умозаключения, уже упоминавшаяся ранее.

Покажем работу правил на тождественно истинном высказывании: $(A \wedge B) \rightarrow B$. Можно экспериментально найти или посмотреть в учебнике по математической логике, чему эквивалентно выражение $A \wedge B$. Выражения с такой же таблицей истинности существуют, например $\neg(\neg A \vee \neg B)$. Здесь знак $\neg$ означает отрицание. Методом замены можно получить следующую теорему:

$\neg(\neg A \vee \neg B) \rightarrow B$.

Переменная B встречается в теореме дважды. Воспользуемся подстановкой и заменим B на $\neg(\neg A \vee \neg B)$. В результате получим следующее выражение:

$$\neg(\neg A \vee \neg\neg(\neg A \vee \neg B)) \rightarrow \neg(\neg A \vee \neg B).$$

В посылке импликации есть двойное отрицание. Отрицание отрицания возвращает выражение к исходному. Получаем таким образом:

$$\neg(\neg A \vee \neg A \vee \neg B) \rightarrow \neg(\neg A \vee \neg B).$$

Внутреннюю скобку мы убрали, так как в посылке две одинаковые операции, отделять которые скобкой не имеет смысла. Рассмотрим часть посылки $\neg A \vee \neg A$. Очевидно, что значение этой дизъюнкции совпадает со значением отрицания. А значит, дизъюнкция эквивалентна входящему в нее отрицанию, и получаем:

$$\neg(\neg A \vee \neg B) \rightarrow \neg(\neg A \vee \neg B).$$

Теперь и в посылке, и в следствии стоит одно и то же выражение, следовательно, мы получили теорему, истинность которой не требует проверки таблицей, так как получена обыкновенная тавтология, из истинного утверждения оно само следует с полной очевидностью.

Эвристические механизмы машины «Логик-теоретик»

Три описанных выше правила дают практически неограниченные возможности вывода теорем в исчислении высказываний. Но, во-первых, как уже было показано на примере, весьма возможен результат, не имеющий смысла, вроде полученной тавтологии. Во-вторых, и это более принципиально, задача получения всех возможных истинных утверждений и не ставится. Более интересно, взяв некое выражение, к которому у исследователя имеется интерес, выяснить, можно ли его вывести из уже известных теорем. Собственно, так наука и работает. Выдвигается гипотеза, а затем либо ищется ее подтверждение, либо опровержение.

Есть простейший алгоритм, позволяющий быстро построить такую программу. Этот алгоритм, состоящий из трех шагов, уже был описан в начале главы. Напомним, его проблема в том, что множество теорем растет очень быстро, превращая дерево перебора теорем

в практически необозримый лес. И единственный выход из положения – разработка эвристик, позволяющих обрубать ветви перебора. В создании таких эвристик и заключается главное достижение авторов программы «Логик-теоретик». Эвристики, применяемые программой, сводятся к нескольким несложным идеям.

Метод отделения. Допустим, необходимо доказать теорему *B*. Пусть в результате анализа выясняется, что теорема *B* легко доказывается, если получится доказать теорему *A*.

В математике, вообще, такие ситуации – дело обычное, вспомогательные теоремы, помогающие доказать исходную, называются леммами. Пример:

Теорема. Сумма углов любого четырехугольника равна 360°.
Лемма. Сумма углов любого треугольника равна 180°.

В школьном курсе равенство суммы углов любого треугольника не доказывается, это сообщается как факт. И мы доказательством леммы заниматься не будем, посмотрим лишь, как ее можно использовать для доказательства исходной теоремы. Для этого достаточно взглянуть на рис. 6.1:

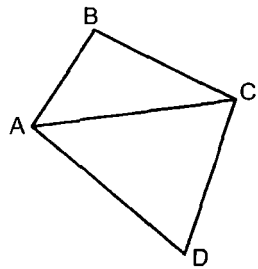

Рис. 6.1 ❖ Пример использования леммы

Видно, что четырехугольник можно поделить на два треугольника, сумма углов каждого из которых равна 180°, а значит, сумма углов всего четырехугольника равна 360°. Следовательно, доказательство леммы непосредственно даст и доказательство теоремы. Примерно так работает принцип. Его реализация в исчислении высказываний в некотором смысле даже легче, чем в геометрии, в силу простоты теории. Если доказываемое выражение есть *B*, метод отделения ищет аксиому или теорему вида «$A \to B$». Если таковая найдена, доказательство *A* считается новой подзадачей. Далее, если можно доказать *A*, то

поскольку «$A \rightarrow B$» – теорема, B также будет доказано. Как в нашем примере. Из леммы теорема следует непосредственно, поэтому если лемму принять как истину или доказать ее, то дело сделано.

Методы цепеобразования. Эти методы используют транзитивность отношения импликации для создания новых подзадач, решение которых дает доказательство исходной теоремы. Из свойства транзитивности напрямую следует, что если исходное утверждение имеет вид «$A \rightarrow C$», то можно попробовать найти теорему вида «$A \rightarrow B$». Если таковая теорема существует, то можно сформулировать подзадачу доказательства теоремы «$B \rightarrow C$». Действительно, по свойству транзитивности, если теорема «$A \rightarrow B$» истинна и теорема «$B \rightarrow C$» также истинна, то исходную теорему «$A \rightarrow C$» в силу определения импликации также необходимо считать истинной. Такая цепь называется прямой. Похожим образом работает и метод обратной цепи. На первом шаге ищется теорема вида «$B \rightarrow C$», и в случае удачи доказательства теорема «$A \rightarrow B$» объявляется новой подзадачей.

Обе цепи – и прямая, и обратная – ровным счетом ничего не гарантируют. В обоих случаях делается предположение, что некую теорему можно доказать, и если доказательство действительно будет получено, то это поможет в поиске доказательства исходной теоремы. Но новая задача, полученная в цепи, не обязана вести к исходной задаче. Более того, новая задача даже не обязана оказаться доказуемой теоремой, что создает очень значительную трудность, выше уже было упомянуто, что если построением дерева теорем крайне трудно дойти до нужной теоремы, то практически невозможно таким способом выяснить недоказуемость утверждения. А это, в свою очередь, означает, что недоказанные и даже недоказуемые утверждения будут непрерывно порождаться в процессе вывода. А если мы вспомним, что ложных утверждений неизмеримо больше, чем истинных, то дело совсем плохо, но, как ни странно, программа «Логик-теоретик» по сообщениям авторов оказалась достаточно успешна.

Базы знаний

Впрочем, успешность программы «Логик-теоретик» мало что доказывает. Исчисление высказываний – слишком простая теория, чтобы по ней можно было судить об успехе метода. Может быть, эксперимент по поиску доказательств был бы более убедительным, если бы появилась программа, работающая хотя бы в исчислении предикатов, но дело даже не в этом. По-настоящему успешный процесс решения

задач должен направляться уже полученным знанием, корректироваться достигнутыми успехами и неудачами. Проанализированная неудача должна давать информации больше, чем просто отсечение ветки вывода утверждений на дереве поиска. Человек, получая отрицательную информацию, не только делает вывод о том, что некие теоремы уже невозможны, он определяет и то, что теперь возможно на фоне этих неудач.

Для анализа человеческого интеллекта, например, очень важно понять, каким образом неудача объяснения опыта Майкельсона–Морли в рамках физики Ньютона могла подвигнуть физиков не только к умозаключению об ограниченности ньютоновой механики, это как раз понятно, но и направить их к созданию релятивисткой физики.

Движение вперед в интеллектуальной деятельности становится возможным потому, что информация, хранящаяся в нашем разуме, представляет собой нечто большее, чем упорядоченное множество истинных или ложных высказываний. Для этого феномена даже был придуман специальный термин – «база знаний».

Появление нового термина само по себе ничего не означает и ничего не дает, необходимо определить его содержание, увязать его с терминами, понимание которых не создает проблем

Приведем несколько примеров, определяющих понятие базы знаний.

База знаний – база данных, содержащая правила вывода и информацию о человеческом опыте и знаниях в некоторой предметной области. В самообучающихся системах база знаний также содержит информацию, являющуюся результатом решения предыдущих задач. Современные базы знаний работают совместно с системами поиска и извлечения информации. Для этого требуются некоторая модель классификации понятий и определенный формат представления знаний. Иерархический способ представления в базе знаний набора понятий и их отношений называется онтологией.

База знаний представляет собой определенную (логическую, алгоритмическую, семантическую, фреймовую, интегральную) модель, предназначенную для представления человеческих знаний в определенной предметной области, – это совокупность моделей и правил, порождающих анализ и выводы, необходимые для нахождения решений сложных задач в некоторой предметной области.

База знаний – семантическая модель, описывающая предметную область и позволяющая отвечать на такие вопросы из этой предметной области, ответы на которые в явном виде не присутствуют в базе.

Заметим, что первые два определения опираются на понятие человеческого знания. Уже одно это говорит об их низком качестве. Фактически здесь понятие знания определяется через само себя, что не имеет смысла. Приставка «человеческое» не спасает положения, так как «человеческое знание» – это, очевидно, частный случай знания вообще, а общее через частное определять нельзя. Логически правильно было бы дать общее определение и уже через него выразить человеческое. Третье определение говорит о том, что позволяет база знаний. С тем же успехом мы могли бы определить самолет как то, на чем можно летать.

Кроме того, все три определения используют не менее сложные понятия, чем определяемое: семантическая модель, предметная область, правила, порождающие анализ, самообучающиеся системы. Сложность и нечеткость представления каждого из этих понятий усиливают проблему определения исходного понятия.

Знание как предложение естественного языка

Говорить о том, что я обладаю каким-то знанием, имеет смысл только в том случае, если я это знание могу сообщить другому человеку и этот другой сможет принять решение о включении моего сообщения в свою базу знаний или аргументированно его отвергнуть. Этот процесс требует посредника. В нашем, человеческом случае посредником является естественный язык. И можно не сомневаться, что в любом случае, для любой интеллектуальной системы должен быть посредник – язык. Средства выражения такого языка не ограничиваются человеческим вариантом, здесь фантазия может предложить самые различные технологии, но некие общие свойства, дающие возможность определить технологию передачи как языковую, должны быть. Отсюда возможна следующая программа определения термина «знание».

- Необходимо определить, что должен представлять собой язык общения интеллектуальных систем. Очевидно, потребуются простейшие смысловые единицы и правила их объединения в правильные предложения.

- Необходим критерий, позволяющий определить осмысленность предложения.
- Необходимо построить критерий, отличающий просто осмысленное предложение от предложения, содержащего знание.
- После этого знанием можно назвать предложение языка.

Такой язык, очевидно, должен принципиально отличаться от формальных алгоритмических языков. В языках, используемых в качестве посредника между машиной и человеком, огромную роль играет набор синтаксических правил. Семантические (смысловые) правила тоже существуют, но они как раз почти на уровне здравого смысла. В естественном языке центр тяжести сильно смещен в сторону семантики. Синтаксис тоже есть, но он как бы не совсем обязателен. Даже очень безграмотный текст вполне можно понять, чего не скажешь о компьютерной программе, где даже одна синтаксическая ошибка делает ее нечитаемой для машины. В естественном языке проблемы смысла разрастаются до огромной величины. Поэтому разработку языка искусственного интеллекта разумно начинать именно с семантики – смысла – знания. И мы пошли по кругу. Мы знание определили как предложение естественного языка. В языке нам потребовалось понятие смысла, а оно опять потребовало понятия «знание». Попробуем другой путь.

Картина мира как конструкция классов, объединяющих родственные объекты

В своей речи человек говорит об объектах и отношениях между ними. Объекту соответствует абстрактное языковое понятие. Поэтому далее под объектом будем понимать нечто наблюдаемое или мыслимое, а под понятием – языковую единицу.

Если человек произносит слово «яблоко», то, возможно, он имеет в виду некое конкретное яблоко, но сказанное слово может означать любое яблоко, в том числе яблоко, про которое человеку ничего не известно. И даже еще интереснее, объект не обязан быть объектом физического мира. Это вполне может быть объект мира интеллекта (будем в дальнейшем называть его внутренним). Например, «Целочисленное уравнение», «Гомеоморфизм», «Валентные связи».

Особенность «увиденного» органами чувств или мышлением объекта – в том, что он существует вне времени, в том смысле, что нет необходимости, указывая на объект, ссылаться на время его наблюде-

ния, и он автоматически увязывается с языковым понятием, которое тоже существует вне времени. Если мы говорим «яблоко» или «уравнение», то, конечно, речь занимает определенный промежуток времени, но этот факт несущественен.

Таким образом, объект можно определить как то, что интеллект способен выделить из наблюдаемого мира (не важно, физического или внутреннего), и акт отделения объекта независим от времени.

Независимость объекта от времени – важный момент, так как есть еще такая вещь, как процесс или действие, включающие время в свое описание, и время нам понадобится, чтобы отделить процесс мышления от знания как результата мышления. Вернемся к объекту. Ему соответствует понятие, выражаемое в языке термином (именем). Следующим шагом требуется понять, как из множества имен создать систему. Это необходимо, поскольку множество терминов, обозначающих объекты, не свалено в одну большую кучу. Поэтому следующий пункт наших рассуждений – о том, что представляет собой система имен и на базе чего она создается.

Прежде всего отметим, что имя – не просто обозначение некоего объекта. Оно необходимо для указания его места, во времени, в пространстве или для указания его родовой связи с другими именами объектов.

Что значит указание местоположения во времени и пространстве, думаю, пояснять нет нужды, а вот что такое «родовая связь», опишем детальнее. Наше мышление, обнаруживая в самом себе различные объекты, встраивает их в иерархию классов подобных объектов. Например, яблоко – это фрукт (входит в класс фруктов), яблоко – это съедобный предмет, яблоко – это круглый предмет. То есть родовая связь указывает на множество объектов, которое содержит похожие с точки зрения какого-то признака и в которое можно включить и данный.

Наш объектно-понятийный мир отражает статичную вселенную. Поэтому вполне достаточно будет придумать метод разбиения объектов на классы и определить правила включения класса в класс и пересечения двух классов. Например, объект класса «Березы» входит в класс «Деревья», а оно, в свою очередь, в класс «Растения», а он, в свою очередь... в общем и т. д. Глубина включения может быть сколь угодно велика. Классы могут пересекаться. Например, объект «Кошка» входит в класс «Кошачьи» и в класс «Домашние животные», но

представитель домашних животных – это не обязательно кошка, а представитель кошачьих – это не обязательно домашнее животное. Эта конструкция называется в теории множеств пересечением.

Создав систему взаимопересекающихся и объемлющих друг друга классов, мы получаем вполне обозримую систему объектов внешнего и внутреннего мира. Правило создания такой системы основано на двух типах объектов: собственно объект, существующий сам по себе, и объект-свойство, не существующий сам по себе. Например, «синий», мы можем выделить такую сущность, и можем ее мыслить, значит, это объект, но он существует только в связке с объектом-носителем: синее небо, синяя краска, синий нос и т. д. Появление объекта-свойства создает возможность образовать класс. Каким образом?

Очень простым и естественным. Свойство, играя роль уточнения для объекта, выделяет подкласс. Съедобный плод есть подкласс плодов. Кубанское яблоко есть подкласс яблок и т. д. Таким образом, имея класс объектов и набор свойств, можно, применяя любую комбинацию свойств, получать новые подклассы. На базе такого с виду простого отношения, как отношение уточнения (будем далее называть его так), возможно, создать сколь угодно сложную систему классов. Можно уточнить пространственное положение объекта – вещь, лежащая левее от меня. Можно временное: телевизор, купленный в прошлом году. Это означает, что пространственные и временные отношения можно рассматривать как частный случай родового отношения уточнения.

Заметим еще раз, что свойство, уточняющее объект внешнего или внутреннего мира, тоже является объектом, лишь более высокого уровня абстракции, а значит, свойства также могут участвовать в таком отношении. Например: «вкусный красный арбуз». Здесь объект «арбуз» уточняется двумя свойствами. А здесь – «красное с желтыми пятнами яблоко» – объект яблоко уточняется сложным свойством, корень которого «красный» уточняется свойством «с желтыми пятнами».

Объектом некоторого уровня абстракции может быть все, что угодно, что в процессе наблюдения может выделить разум. Например, взаимное положение объектов: «Расстояние между этими двумя деревьями примерно километр». Мы не обязаны создавать понятие расстояния для каждой пары деревьев, а значит, расстояние становится самостоятельной сущностью. И т. д. Все, что мы можем выделить в процессе наблюдения, может стать объектом мышления, любой объект может, участвуя в отношении уточнения другого объекта,

становиться его свойством. Эти две ипостаси: объект – свойство, – присущие любой наблюдаемой сущности, заменяя друг друга, в зависимости от места сущности в отношении уточнения создают всю наблюдаемую, статическую (именно статическую, это важное дополнение) картину мира *любой сложности*.

Для того чтобы отразить этот интеллектуальный механизм в языковой форме, необходимо правило преобразования имени объекта в имя-свойство. В естественных языках такой механизм есть. Например, в русском языке имена существительные легко превращаются в имена прилагательные. Обратное не всегда возможно. Например, свойство круглый не существует само по себе в виде чистого объекта (имени существительного). Цвет, геометрическая форма, вес и многое другое не существуют сами по себе. Но преобразование не обязано быть всегда обратимым. Еще раз сформулируем главную идею:

С целью создать из множества наблюдаемых сущностей упорядоченную картину мира мы наделяем их способностью быть как объектами, так и свойствами объектов. Свойства, вступая с объектами в отношение уточнения, разбивают все множество мыслимых и наблюдаемых объектов на систему взаимовложенных и взаимопересекающихся классов. Полная система классов, в которые входит объект, исчерпывающе описывает его положение в мире – и пространственное, и временное, и родовое.

Но такая картина мира статична. Процесс в ней также можно определить как объект, обладающий рядом свойств и поэтому относящийся к некоторому множеству классов. Так устроено наше знание. Оно действительно статично. Утверждая, что два тела притягиваются друг к другу прямо пропорционально произведению масс и обратно пропорционально квадрату расстояния (Всемирный закон тяготения), мы объявляем некий факт, уже существующий за пределами времени и не представляющий собой процесса. Само явление тяготения – это процесс. Но все, что мы о нем можем утверждать, имеет вид статического, завершенного знания. Сказанное выше дает возможность дать еще одно определение базы знания:

База знаний – это множество понятий, увязанных в систему отношением уточнения

Не буду утверждать, что это определение свободно от недостатков, но у него есть одно преимущество. Оно определяет термин «знание» через простые и хорошо объясняемые вещи.

Операционное знание

Однако завершенное знание, хранящееся в базе знаний, – это не все, что мы можем обнаружить в мышлении. Определим еще одну интересную сущность, которую в дальнейшем назовем операционным знанием.

Этот термин будем использовать для обозначения содержания мыслительного процесса. Нечто такое действительно существует, так как мышление, очевидно, не сводится к воспроизведению или комбинации знания, выражаемого текстом, написанным на естественном языке. И даже более того, наш внутренний язык, которым мы пользуемся для сопровождения мыслительного процесса, существенно отличается от текстов, которые мы пишем. Бумажные и электронные тексты также не отражают процесса, это все результат мыслительного действия.

Прежде чем более четко выразить, что такое знание операционное и какова его роль в мыслительном процессе, необходимо еще немного поговорить о том, что такое завершенное знание. Его определение как знания, записанного в книге (электронной или бумажной), нас не может устраивать в силу слабой информативности. Утверждать, что книжное знание – это верное знание, в то время как мыслительный процесс оперирует гипотезами, тоже ошибочно. Гипотеза также может быть книжным, завершенным знанием. Две тысячи лет аксиома параллельных в геометрии несла в себе признаки гипотезы, но ей активно пользовались. Великая теорема Ферма сотни лет находилась в статусе гипотезы, гипотезы тысячелетия сформулировал Пуанкаре. Вообще, гипотеза – весьма распространенная форма знания. И это не оговорка, гипотеза – тоже знание, но знание о возможном.

У завершенного знания есть несколько определяющих свойств. Во-первых, в отношении его системы можно сказать, что она (система) непротиворечива. Это означает, что в системе могут сосуществовать только такие истинные утверждения, использование которых в мыслительном процессе не может привести к ложному утверждению. В такой системе могут существовать утверждения в ранге гипотез, причем гипотезы могут друг другу противоречить, но это лишь гипотезы.

Второе определяющее свойство завершенного знания – его полнота. Это означает следующее: любой вопрос, заданный в терминах системы и относящийся к предмету системы (в рамках квантовой механики нельзя задавать вопрос о способе размножения вирусов), имеет

ответ в рамках этой же системы. И наконец, третье: утверждения завершенной системы безличны. Они не соотносятся ни с какой-либо задачей, ни с каким-либо исследователем. Пояснim. Высказывание «Пифагор полагает, что сумма квадратов катетов прямоугольного треугольника равна квадрату гипотенузы» имеет личностное содержание. Это не просто гипотеза, это гипотеза, связываемая с конкретным ученым. То же утверждение в школьном учебнике с исторической ссылкой, что его впервые сформулировал древний грек, – уже завершенное знание. Утверждение «Я не верю в корпускулярную природу света» личностное. Но и общепринятое: «Я полагаю, что природа света дуалистична, это одновременно и волна, и поток корпускул» – тоже личностное. И наконец, завершенное знание статично, его формулировка твердо определена, и если формулировок существует несколько, то они идентичны, как идентичны словесная формулировка теоремы и формула, кратко ее выражающая.

Свойства завершенной системы знания можно детализировать в рамках отдельной книги, но мы остановимся на этих четырех, лично мне они кажутся определяющими, и на них можно показать отличие операционного знания от завершенного.

Операционное знание не находится вне времени и вне процесса. Оно динамично. Формулировки используемых утверждений могут меняться в процессе мышления. В рамках операционного процесса я могу сказать: «Квадрат – это четырехугольник, противоположные стороны которого равны». Я имею на это право, но когда я обнаружу контрпример (четырехугольник с указанным свойством, не являющийся квадратом), я смогу изменить текст этого утверждения. Еще один важный момент, который можно отнести сюда же. Объекты завершенного знания существуют в виде хорошо определяемых языковых понятий, объекты операционного знания существуют в виде образов, и это еще один механизм, обеспечивающий динамизм системы. А главный механизм развития заключается в существовании противоречивых, конкурирующих утверждений. Их существование подвигает исследователя к выбору, для чего нужны веские основания, которые можно получить только в мыслительном процессе.

Операционное знание всегда неполно. Всегда есть вопросы, ответов на которые в рамках операционного знания нет. И это тоже является движущей силой расширения и усложнения структуры знания.

Операционное знание не существует за пределами личности. Оно глубоко индивидуально, зависит от состояния конкретного интел-

лекта, истории его развития, предпочтений, сформированных решаемой задачей и ролью этой задачи в общей мыследеятельности.

Конечно, операционное знание и завершенное – это крайние точки мыслительной деятельности. В любой завершенной теории есть неясные моменты, что может означать существование скрытых и пока не видимых противоречий. Завершенное знание, даже школьные и вузовские учебники, нельзя считать полностью безличным, хотя его содержание и определяется общепринятой точкой зрения. Это действительно так, но надо понимать, что так называемая «общепринятая точка зрения» – на самом деле не более чем точка зрения группы людей, может быть, очень большой группы, может быть, эти люди очень влиятельны в настоящее время, но эта точка зрения всегда «не от бога».

В то же время и операционное знание не представляет собой абсолютного хаоса. Во-первых, в нем всегда используются завершенные утверждения, изменять смысл и формулировки которых человек в здравой памяти и трезвом уме не будет. Любой мыслящий человек в своем мышлении старается двигаться в сторону обособленности от личного. В общем, система операционного знания всегда стремится превратиться в завершенную, ей это никогда не удается, поэтому мыслительный процесс есть процесс вечный и непрерывный.

Самоорганизующийся искусственный интеллект

Ядром разума является не завершенное, формализованное знание, а живое операционное. А значит, вопрос заключается в том, как построить и заставить работать эту форму существования знаний. Для формулировки ответа нам потребуется понятие самоорганизующейся системы. Система вообще – это набор элементов, участвующих в общем процессе, например детали автомобиля создают систему, способную двигаться. Особенность такого механизма – потребность во внешнем управлении для поддержания своей функциональности. Есть другие системы, способные поддерживать работоспособность собственными силами, если не всегда, то, по крайней мере, некоторое время, пока внешние силы не окажут на систему слишком сильного влияния.

Такие системы называются самоорганизующимися. Простейший пример – волчок. Если момент вращения достаточно велик, то не-

большое отклонение от вертикальной оси не приводит к его падению, сразу же после воздействия у волчка возникает собственная сила, компенсирующая внешнюю. Еще один пример такого рода, но более сложный – двухкомпонентная экосистема хищник–жертва. Если хищников становится слишком много, это приводит к уменьшению популяции жертвы, кормовая база хищников сокращается, что ведет уже к уменьшению популяции хищников. Как результат численность обеих популяций в течение времени испытывает колебания около некоего положения равновесия. Такого рода схема самоорганизации характеризуется равновесием между действием и противодействием. Противодействие возникает неотвратимо, по природе системы.

Более сложная схема держится на принципе конкуренции, в рамках которого разные элементы системы борются за использование ограниченных ресурсов. А еще более сложный принцип сотрудничества элементов системы предполагает совместные действия на общую пользу. Такая совместная деятельность участников самоорганизующейся системы совершенно не обязательно строится на базе разума. Например, на нашей планете очень широко распространено сотрудничество между растениями и насекомыми. Растения дают насекомым пищу в виде нектара, взамен насекомые, и не только пчелы, оказывают услуги по опылению. Действия обеих сторон в этом механизме совершено неразумны. Из примеров видно, что механизмы самоорганизации создают системы, способные развиваться и приспосабливаться к изменяющимся условиям без внешнего управления.

Посмотрим, каким образом понятие самоорганизующейся системы можно использовать для построения саморазвивающегося разума. Определим цель работы разума как непрерывное расширение базы знаний. Это не совсем точно, интеллект не сводится к знаниям, есть еще методы получения знаний, но методы и знания жестко отделить друг от друга нельзя, поэтому в некотором приближении можно интеллект и базу знаний считать вещами идентичными. В последней главе мы еще вернемся к этой проблеме. А пока так: цель разума – непрерывное саморасширение, то есть увеличение объема знаний.

Выше мы уже определили базу знаний как множество понятий, связанных отношениями. Уже было сказано об особой роли отношения уточнения, даже утверждалось, что на его основе можно определить все отношения, возможные между понятиями. Можно согласиться и с таким скупым подходом. Можно определить некоторое количество более специализированных отношений, например отношения пространственного положения: ближе, дальше, слева, справа и т. д.

Такое право на существование имеют отношения временные: раньше, позже, одновременно и т. д. Существенно значимо здесь то, что отношения устанавливают на множестве понятий порядок в самом общем смысле этого слова. Линейный порядок – простейшая форма порядка вообще, устанавливается, например, временными отношениями. Заниматься формальным определением отношений не будем, надеюсь, интуитивно смысл термина понятен. Отношения и понятия создают знания, выражаемые в языке истинными утверждениями.

Отдельно выделим отношение включения. Оно определяет структуру базы знаний, ее разделение на относительно замкнутые области знания. Например, понятие «наука физика» включает в себя все понятия, относящиеся к описанию материального мира, и создает одноименную область знания. Множество понятий, структурированное отношениями, создает статическую структуру знания. А теперь попробуем сделать небольшую добавку, из специй превращающую эту структуру в развивающееся операционное знание.

Введем понятие «идеи». Идея – это вопрос к базе знаний и гипотеза о направлении поиска ответа. Техника постановки вопроса может сводиться к добавлению к утверждению какой-либо вопросительной формы. Например: «Алгебраическое уравнение может иметь целые корни», это утверждение можно прямо переформулировать в вопрос, используя множественную форму словосочетания «целые корни»: «Сколько целых корней может иметь алгебраическое уравнение?» Или используя неопределенность отношения, введенного словом «может»: «Каково условие наличия у алгебраического уравнения целых корней?» Или используя знание о том, что одна из целей алгебры состоит в том, чтобы давать методы решения задач: «Как найти целые корни алгебраического уравнения?» Техника формулировки вопроса у развитого интеллекта может быть очень сложной. Но еще интереснее техника формулировки гипотезы.

Сразу заметим, что речь идет не о гипотезе решения, а о гипотезе направления поиска решения. Это необходимый предварительный шаг. Функция гипотезы – в определении области знания, в которой будет выполняться поиск решения. Это означает, что гипотеза должна задать множество понятий. Эти понятия подтянут понятия, связанные с ними через отношения. Автоматически в эту область войдут истинные утверждения, построенные на выбранных понятиях и связанных с ними отношениях. Проиллюстрируем механизм на примере с алгебраическим уравнением. В вопросе упоминается такой объект, как целое число. Мы знаем, что есть понятие целого числа и область

знания, описывающая поведение целых чисел. С целым числом связано понятие делимости. Это создает возможность искать решение, используя понятие делителя. Делитель целого также является целым числом, значит, на этапе поиска решения можно попытаться определить все, что касается целых корней, через делители коэффициентов уравнения (это единственные целые числа в уравнении), и как мы знаем, действительно, целочисленные корни находятся среди делителей свободного члена. А значит, исходная гипотеза, принявшая в качестве области поиска решения теорию целых чисел, была полезной.

Идея является движущей силой интеллекта. Разум, генерируя и отрабатывая идеи, получает новые знания, каковые становятся отправной точкой для формулировки новых идей. А значит, идея и есть главный инструмент самоорганизующейся системы интеллекта.

Идеи могут возникать совершенно свободно и хаотично в процессе развития базы знаний, однако в силу этой самой свободы развитие двух идей может привести к противоречию между собой. Две противоречивые идеи могут сосуществовать в операционном знании, но не бесконечно долго, так как противоречивое знание не может быть сохранено как завершенный результат. Это означает, что для поддержания разума в состоянии равновесия или, что еще лучше, в состоянии развития необходимо выработать механизм конкуренции идей.

Этот вопрос никак нельзя посчитать чисто техническим. Механизм конкуренции имеет глубоко принципиальное значение, определяя эффективность разума, и, наверное, он должен быть достаточно сложен. Но в первом приближении возможны и простые решения. Например, можно посчитать объем знания, даваемый противоречивыми идеями, и объем знания, ими разрушаемого, и принять ту идею, которая обещает максимальный выигрыш. Например, специальная теория относительности объясняет все, что объясняет механика Ньютона, но она дает и нечто большее. Такой подход очень прямолинеен и примитивен, но, думаю, способен дать развивающийся разум.

Последнее замечание

В этой главе вопросов больше, чем ответов. Но что поделаешь, сложность темы обязывает. А в заключение попробуем найти корни еще одного важного феномена разума – интуиции. Этот феномен отличается тем, что, ощутив интуитивное озарение, мы не знаем, откуда пришло решение. Конечно, мы много о нем думали, мы уже много знаем, но все же четко ответить на этот вопрос, откуда решение при-

шло, не можем. Решение проблемы, на мой взгляд, лежит в области психологии разума. Есть часть мышления, не контролируемая сознанием, разум не рефлексирует на всем, что в нем происходит. Позволю себе предположить возможный механизм интуиции. Выше говорилось о том, что «идеи» генерируются разумом совершенно свободно и хаотично. Но развивать их все невозможно, у разума не хватит материальных ресурсов. И появляется фильтр – рефлексия – и так называемый здравый смысл. Здравый смысл – это эвристический алгоритм определения эффективности идеи до ее проверки или после минимальной проверки. Возможен, например, такой алгоритм (критерий): некоторые знания имеют статус догмы (непререкаемого знания), если идея противоречит догме, то она признается противоречащей здравому смыслу и отсеивается.

Понятно, что таким образом отсеиваются все возможности для создания новых теорий, так, плоская земля, стоящая на трех слонах, а те, в свою очередь, на черепахе, тоже была догмой. Проблема получения радикально нового знания лежит, как ни странно, в ограниченности сознания. Разум не в силах бороться с догмами, но часть мыслительного процесса уходит на уровень подсознания, не контролируемого рефлексией. И именно там, на подсознательном (или сверхсознательном, кому как нравится) уровне, могут развиваться самые сумасшедшие идеи. Рано или поздно они, как и любая осознанная идея, либо погибают в конкурентной борьбе, которая, кстати, тоже может быть подсознательным механизмом, либо выходят на уровень сознания, обретя достаточно мощную доказательную базу, способную преодолеть догму. Это воспринимается сознанием как интуитивный прорыв. А на самом деле он обеспечивается богатой идейной составляющей и низким уровнем барьера догм у конкретного человека.

Я не буду утверждать, что предложенный механизм действительно работает в случае человеческого разума, но полагаю, что он вполне может быть реализован для разума искусственного, а именно ему и посвящена эта книга.

Глава 7

Интеллект, равный человеческому?!

Не просто техническая проблема

Моделирование разумного поведения в конкретной ситуации, виде деятельности – действительно очень интересная задача, обещающая, да уже и давшая, человечеству новые технологии и потрясающие возможности. Но это не означает, что задача создания интеллекта, равного человеческому или даже превышающего его, не осталась на повестке дня. Эта задача одновременно очень интересна и очень опасна. Если полноценный искусственный интеллект возможен, то, скорее всего, он будет мощнее человеческого, хотя бы в силу практически не ограниченной памяти и намного более высокой скорости обработки данных, к которым присоединятся наше творческое мышление и способность к самосознанию, сюда же можно будет приплюсовать возможность прямого общения искусственных разумов сетевыми средствами. Если это случится, если существа, обладающие таким интеллектом, появятся, то, возможно, мы для них будем примерно тем же, чем для нас являются приматы. Искусственный интеллект может стать реальным началом конца человечества, по той простой причине, что контролировать существ, превышающих нас по развитию, мы уже не сможем, как, например, мы можем контролировать ядерное оружие. Но так устроена человеческая наука – если что-то возможно, то она будет к этому стремиться.

А теперь главный вопрос: так можно ли создать полноценный искусственный интеллект, способный развиваться самостоятельно? Мне так кажется, что ответ на этот вопрос очень прост. Если принять версию божественного происхождения человека, то нет, стать творцом, равным богу, человеку не удастся никогда, слишком велика качественная

разница. А если человек и его разум – продукт эволюции, то в основе нашего разума лежат воспроизводимые механизмы, и тогда все решает время. Тогда вопроса «Возможно это или нет?» не существует, а есть вопрос «Когда это случится?».

Но как понять, что полноценный искусственный интеллект уже существует? Любопытно, что наиболее простой и в то же время практически применимый критерий возник примерно тогда же, когда была поставлена и задача, я имею в виду тест Тьюринга. Вспомним его суть. Если человек, беседуя с машиной, не сможет понять, что перед ним машина, то, значит, он беседует с разумным существом. Идея Тьюринга вполне понятна, разумная речь считается безусловным свойством разума.

В тесте Тьюринга есть, однако, слабое место. Быть свойством разума и быть эквивалентом разума – это не одно и то же. Конечно, любое мыслящее существо (не только человек) просто обязано обладать знаковой системой передачи информации, мы такую систему называем речью. Но речь – это форма выражения результата мышления, а не само мышление. Это означает, что в мышлении обязательно есть механизмы, речью не выражаемые, не заключенные в речи. Например, интуиция. Совершенно не понятно, какую роль в интуитивном прозрении может играть текст, написанный на бумаге или произнесенный голосом. Например, образное мышление, которое есть у каждого человека, не использует слово. А если речь не эквивалентна мышлению, то не значит ли это, что разумную речь можно промоделировать без участия разума? Еще первые разработки в этой области, программы «Доктор» и «Элиза», показали, что это действительно возможно.

Человеческий язык, любой: русский, английский, суахили и т. д. – представляет собой набор правил, позволяющих алгоритмически строить синтаксически правильные предложения. Этих правил достаточно много, особенно если считать исключения (собственно, это тоже правила), но их набор конечен, и для современных быстродействующих компьютеров обработка синтаксиса проблемы не составляет. Осмысленность предложений – проблема, решаемая объемом базы знаний. Если она достаточно велика, то, может быть, с точки зрения такого продвинутого компьютера человек окажется не вполне разумным, так как у большинства людей база знаний очень ограничена. Более того, технически не сложно вложить в память компьютера объем знаний, гарантированно превышающий базу знаний любого представителя человеческой расы, что, однако, не отменяет разумности человека и не добавляет разума машине.

Таким образом, тест Тьюринга, требующий лишь продемонстрировать разумную речь, вряд ли можно считать достаточным критерием. Наверное, такая машина должна показать не общие разговорные навыки, а что-то специфическое, включающее в себя глубинные интеллектуальные механизмы.

На мой взгляд, человек отличается от говорящей машины тремя вещами. Во-первых, он способен принимать решения, и рамок, ограничивающих эту возможность, нет. Если, к примеру, дано задание решить уравнение, то человек может выбрать тот или иной метод, но может отказаться от решения уравнения вообще. Если человек идет в магазин с конкретной целью купить хлеба и колбасы, это еще не означает, что он выйдет оттуда именно с хлебом и колбасой, это даже не означает, что он вообще зайдет в магазин. Цель может на ходу поменяться под влиянием неожиданного внешнего фактора или какой-то сиюминутной мысли.

Во-вторых, человек способен наращивать свою собственную базу знаний, и делает он это не так, как машина. Современный искусственный интеллект, по большому счету, получает готовые знания, и его функция – лишь грамотно упаковать полученное в уже имеющуюся систему. Человек же на вход получает не знания, а информацию – что-то вроде руды, которую еще надо переплавить в знания. Есть, правда, нейронные сети, способные обучаться на примерах, но это обучение больше похоже на выработку шаблона для рефлекторного реагирования, чем на интеллектуальную работу. Вряд ли то, что они умеют, сравнимо с обучением мыслящего существа.

И наконец, самое сложное и самое интересное. Человек способен к рефлексии, то есть к осознанию себя и того, что он делает. Чтобы понять, насколько это важно, подумайте вот над какой ситуацией. Пусть есть некий компьютер, настолько мастерски играющий в шахматы, что ему не могут противостоять лучшие шахматисты планеты. Но даже в этом случае между гроссмейстером-человеком и гроссмейстером-машиной есть принципиальная разница. А именно: человек понимает, что он играет в шахматы, а понимает ли это машина? Вот вещи, которые очень важны для понимания интеллекта, но не укладываются в речевые механизмы.

Можно ли улучшить тест Тьюринга

Ясно, что беседа между мыслящим существом и предполагаемо мыслящим не должна быть беседой общего характера. Задача эксперимен-

татора – поставить перед испытуемым какую-то задачу и проверить ход ее решения. Попробуем развить эту мысль. Само по себе решение какой-либо задачи не есть интересная цель. Если наш испытуемый имеет развитую систему речи и хорошую базу в части доказательства геометрических теорем, к которой добавлен мощный эвристический алгоритм поиска доказательств, то, вполне возможно, он сможет доказать теорему и рассказать об этом. А если он и не сможет достичь такой цели, то неудача не будет опровержением его интеллектуальности. Поэтому, во-первых, необходимо контролировать решение «неизвестной задачи», а во-вторых, по большому счету, интересен не результат, которого может и не быть, а лишь процесс.

В критерии Тьюринга участвуют двое: экспериментатор, в интеллектуальности которого нет сомнений, и испытуемый, интеллектуальность которого подлежит установлению. На их беседу не накладывается никаких ограничений, в том числе и временных. Усложним схему. Теперь в испытаниях участвуют трое. ***Экспериментатор***, обладающий интеллектом, и ***Испытуемый***, интеллектуальность которого проверяется. ***Экспериментатор*** и ***Испытуемый*** не знают, что один из них – машина. Третий участник – ***Наблюдатель***. Он человек. ***Экспериментатор*** и ***Испытуемый*** не знают о существовании ***Наблюдателя***. ***Наблюдатель*** получает всю возможную информацию о процессе общения двух других участников процесса, не имея возможности вмешиваться. ***Наблюдатель*** выносит вердикт не относительно интеллектуальности машины, ему даже не обязательно знать, что один из участников – машина. Он делает заключение об их равенстве или неравенстве. Термин «равенство» нуждается в уточнении, но об этом чуть позже. Пока заметьте, ни один из участников процесса не обладает полной информацией о партнерах. Каждый что-то знает о себе одном и лишь со своей точки зрения. То есть машина может и не знать, что она машина. Некоторое преимущество, впрочем, небольшое, имеет ***Наблюдатель***. Он знает, что остальные двое есть: ***Испытуемый*** и ***Экспериментатор*** – но не знает, кто есть кто.

Теперь о содержании беседы ***Испытуемого*** и ***Экспериментатора***. Перед ними ставится цель решить некую задачу. Чем является эта задача, не важно. Можно потребовать от них научиться играть в шахматы, можно потребовать решить математическую задачу высокой степени сложности, например найти доказательство великой теоремы Ферма. Задача должна быть такой, чтобы с общечеловеческой точки зрения она считалась творческой и очень сложной. Обязательное условие: ***Экспериментатору*** и ***Испытуемому*** решение задачи

неизвестно, неизвестно вплоть до незнания формулировки. Экспертное знание о задаче есть у ***Наблюдателя***. ***Испытуемому*** и ***Экспериментатору*** представляется любая информация, которую они могут запросить. Они решают задачу совместно, имея возможность контактировать друг с другом совершенно свободно.

Таким образом, в исследовательской деятельности ***Экспериментатору*** и ***Испытуемому*** придется включать неречевые механизмы мышления, что является предметом обнаружения. Кроме того, им придется принимать решения о направлении исследования, решения либо согласованные, либо личные, оба могут вести исследование самостоятельно, лишь ставя друг друга в известность. А теперь важное замечание о критерии равенства. Два человека не обязательно равноправны в плане исследовательского потенциала, не обязательно это и для нашей пары: ***Экспериментатора*** и ***Испытуемого***. Поэтому задача ***Наблюдателя*** – заметить, верно ли, что оба участника пары оказываются полезными, то есть оба способны генерировать идеи, принимать решения, участвовать в принятии общего решения. Полезность обоих и будет означать равноценность. Кто, как говорится, круче, не является предметом теста.

Проблема номер один

Несколько более сложную задачу составляет обнаружение способности к рефлексии. По большому счету, рефлексия, она же самосознание, – вещь внутренняя. Образно говоря, нужен эксперимент, позволяющий определить – понимает ли шахматная машина, что она играет в шахматы.

Начнем с частной функции самосознания – самооценки. Вопрос – не все ли равно шахматисту-человеку с кем играть? Ответ – нет, не все равно. По разным мотивам: амбиция, желание интеллектуального роста, шахматная карьера, простое стремление к интересной игре заставляет человека-шахматиста выбирать себе партнера, игра с которым поможет в удовлетворении мотивов. Смею утверждать, что стремление к развитию – это базовое свойство интеллекта. Самооценка позволяет обнаружить достижение некоторого уровня и установить новый ориентир. То есть нерефлексирующая машина не имеет критериев для выбора партнера, машина с шахматным самоосознанием будет стремиться к интересному партнеру. Будет ли это означать, что такая машина понимает, чем она занимается, играя в шахматы? Наверное, ответ опять отрицательный, по той простой причине, что

такого рода рефлексию можно запрограммировать достаточно простым алгоритмом, а интуитивно понятно, что у этой проблемы не может быть простого решения.

Теперь возьмем машину из предыдущего эксперимента. Мы не знаем, как она устроена, но знаем о ее способности обучаться на решение произвольной задачи. Дадим ей задачу обучиться шахматам. Ясно, что никакого специального алгоритма по выбору игроков у нее нет, следовательно, если она после обучения начнет выбирать себе игроков в соответствии со своим опытом и растущей силой, то, значит, у нее есть способность к самооценке.

Следующим шагом предложим машине освоить еще пару областей человеческого знания. По завершении учебного процесса предложим ей осмысленную шахматную позицию. Напомню, что в нее не заложено никаких шахматных программ, и в тесте на интеллект мы выяснили, что обучаться машина способна. Так вот, если, глядя на шахматную позицию, она сообщит, что это шахматы, и сможет дать анализ ситуации, то, следовательно, она перешла на новый этап и может сообщить: это шахматы, это задача из области физики и т. д. Это уже означает, что машина научилась распознавать интеллектуальные задачи, и приступаем к последнему этапу: Я ИГРАЮ В ШАХМАТЫ.

Я – высшая степень самосознания, означающая, что мыслящее существо может отделить себя от внешнего мира и от выполняемого процесса. Факт осознания означает, что в памяти интеллекта существует знание, не относящееся ни к одному внешнему по отношению к существу объекту. Внешние объекты – это не только материальные, но и вся совокупность абстрактных понятий, выработанных интеллектом. Знание своего Я появляется при реализации мыслительных задач, направленных на себя. Я утверждаю таким образом, что сознание – не прирожденное свойство человека, оно в нас не заложено, а выработано способностью разума к свободному выбору объектов исследования, способностью выбора в качестве объекта своих собственных мыслительных процессов. Вот это событие и означает появление Я, пока слабого, но растущего с каждым новым самоисследованием. Я – это продукт личной эволюции, возникающий неизбежно в полноценном разуме (свободном, самообучаемом, оценивающем).

Минимальный разум

Искусственный интеллект, используемый в современных технических системах, должен уметь решать сложные задачи за ограниченное

время и делать это сразу, так сказать, после нажатия кнопки включения. Это как родить сразу взрослого человека с дипломом о высшем специальном образовании и опытом работы. Такие программы, имитирующие разум, будут очень сложны и потребуют больших баз знаний. Наверное, можно создать универсальный разум, равноценный человеческому, с момента включения, но было бы более интересно создать разум с потенциалом саморазвития, то есть таким, как создала человека природа. Пофантазируем немного, как можно было бы получить такое решение.

Естественно, нас сейчас не интересует элементная база, на каких транзисторах, микросхемах или квантовых элементах будет работать синтетический мозг. Подумаем лишь о теоретических предпосылках и основаниях модели.

Ясно, что в момент включения синтетический мозг не может быть чистым листом бумаги, в таком случае было бы не понятно, какая движущая сила заставит его развиваться, если от внешнего управления (загрузки необходимых программ) мы отказались. Должны быть какие-то минимальные познания и минимальные методы обработки получаемой информации. Собственно, и человеческий мозг вряд ли чист в момент включения, во-первых, есть генетический канал передачи знания, а во-вторых, развивающийся человек в течение 9 месяцев до рождения имеет возможность получать информацию напрямую от мозга его матери.

Итак, на минимальное знание и минимальные методы мы согласны. Но развитый интеллект демонстрирует не только большую базу знаний, он владеет и развитыми методами мышления, а кроме того, знание не просто становится велико по объему, оно растет и качественно. Как это возможно? Каков механизм? Вопрос не праздный. Представьте себе, что есть помещение, в которое подается деревянный брус (информация), и стоит станок (минимальный метод), способный вытачивать ножку для стула. Согласитесь, что нет никаких причин ожидать, что через некоторое время в этом помещении появится что-то, отличное от ножки стула. Чтобы такое невероятное событие произошло, станок должен уметь вытачивать новые детали для самого себя, усложняющие его собственную конструкцию.

Для механического устройства это нереально. А для интеллектуального есть два механизма, создающих практически не ограниченные возможности. Во-первых, имеющееся знание может стать отправным пунктом для получения нового знания. Например, зная, что сумма углов любого треугольника равна 180°, мы можем начать исследова-

ние суммы углов любого многоугольника, и в ряде случаев имеющееся знание станет решающим фактором. Зная свойства одиночного электрического заряда, мы можем начать исследование свойств тока. И т. д. Такие ситуации обычны, примеров бесконечно много, но есть еще более интересная возможность.

Вырабатываемое знание может становиться строительным материалом для методов мышления. Вот как это может быть. Пусть каким-то способом выяснено, что утверждение *A* относительно объекта, обладающего свойством *B*, истинно (то есть это теорема, закон). Затем в результате исследования выясняется, что это же утверждение *A* истинно и для множества иных объектов, обладающих тем же свойством *B*. Из этого события следует новое знание вида: если объект обладает свойством *B*, то утверждение *A* будет истинным. Таким образом, мы получили не просто новое знание, а качественно новое, обладающее свойством общности. Но это же событие дает возможность построения общемыслительного метода, утверждающего, что для объектов, обладающих общими свойствами, возможны общие истинные утверждения, а это уже метод мышления, называемый обобщением.

Конечно, сказанное выше – не более чем набросок, книга не претендует на научную монографию, но, надеюсь, набросок, достаточный для понимания того факта, что на базе минимального знания и минимальных методов мышления возможно развивать неограниченно мощный интеллект. А детальная разработка таких механизмов – конечно, дело очень непростое.

В заключение хочу поразмышлять еще об одной интересной вещи. В обсуждении проблем искусственного интеллекта и сравнении его с человеческим принято говорить о равном или более мощном разуме, но, на мой взгляд, интереснее создавать интеллект не более слабый или более сильный, а построенный на иных принципах. Соображение очень простое. Имея исходное знание и некий набор простых методов, разум после включения начнет развиваться, исходя из возможностей, заложенных этим примитивом, и характера получаемой информации. Внешняя информация играет роль воспитателя разума, а исходный примитив – роль генетического фактора. А теперь представьте себе два начальных разума, получающих после включения одну и ту же информацию, но в них заложены два разных примитива. Надо полагать, что их эволюционные пути будут разными. И вот здесь поле для архитектурных решений поистине безграничное, единственно, надо решить еще одну фундаментальную проблему – определения свойств

примитива, достаточных для запуска индивидуальной эволюции. Но на этом позвольте полет фантазии завершить.

Экспертное мнение

Значимость проблемы искусственного интеллекта для человечества трудно переоценить. Поэтому практически сразу после появления первых ЭВМ значительные интеллектуальные силы были брошены на разработку думающих программ. Энтузиазм был довольно велик, вдохновение разработчиков подпитывалось серьезными успехами и обилием новых интересных идей. Но в процессе развития теории оказалось, что все не так просто и даже, собственно, что такое интеллект, не вполне ясно. Надо было понять, а что на самом деле мы разрабатываем, к чему стремимся и что это даст человечеству и человеческой науке.

В последующих текстах не будет никаких суждений от автора и никакого авторского анализа. Ниже вы познакомитесь с идеями и мыслями тех людей, которые оказали существенное влияние на создание и развитие теории и технологий искусственного интеллекта. Набор приведенных цитат не представляет собой какой-либо системы. Эти тексты – просто законченные содержательные мысли людей, очень хорошо понимающих проблематику искусственного интеллекта.

«Утверждение, что вычислительные машины могут делать только то, что нами в них запрограммировано, интуитивно очевидно и несомненно правильно, но не доказывает ни одного из тех выводов, которые обычно из этого делают.

Человек может мыслить, учиться и творить благодаря тому, что программа, обусловленная его биологической природой, а также те изменения в этой программе, которые вызваны взаимодействием организма с окружающей средой после его рождения, делают его способным мыслить, учиться и творить. Если окажется, что машина мыслит, учится и творит, то этими качествами ее наделила программа. Ясно, что эта программа не в большей степени, чем человеческая, будет требовать существенно стереотипного и повторяющегося поведения, не зависящего от стимулов из окружающей среды и выполняемой задачи. Это будет программа, делающая поведение системы в сильной степени зависимым от внешней среды, от целей поставленной задачи и от сигналов о том, насколько система продвинулась на пути к достижению цели. Такая программа будет каким-то способом анализи-

ровать собственное поведение, обнаруживать свои ошибки и вносить необходимые изменения, которые позволят в будущем повысить ее эффективность» (Герберт Александер Саймон, нобелевский лауреат, член Национальной академии наук США, профессор компьютерных наук и психологии в университете Карнеги-Мелона).

«С самого начала следует четко разграничить два общих подхода к проблеме обучения машин. Первый из них, который можно назвать подходом с точки зрения нейронных сетей, рассматривает возможность создания «обученного» поведения в случайно связанной переключательной сети (или ее модели на цифровой вычислительной машине) в результате применения некоторой системы поощрений и наказаний. Второй, значительно более эффективный подход – создание эквивалента высокоорганизованной системы. Первый подход должен привести к развитию универсальных обучающихся машин. Сравнение размеров переключательных сетей, которые в настоящее время удается практически сконструировать или промоделировать, с размерами нейронных сетей живых организмов свидетельствует о том, что нам предстоит еще пройти большой путь до получения практически пригодных устройств. Второй подход требует составления новой программы для каждого нового приложения, но его можно осуществить уже в настоящее время» (А. Сэмюэль, инженер корпорации IBM, автор программы, играющей против человека в шашки).

«Причина того, что задачи являются действительно "задачами", лежит в том, что первоначальное множество возможных решений, которое дается решающему задачу, может быть очень большим. Действительные же решения могут быть равномерно и редко распределены по этому множеству, а получение и испытание каждого нового элемента может требовать больших усилий. Таким образом, человеку, решающему задачу, фактически не дается множество возможных решений. Вместо этого ему сообщается некая процедура, позволяющая ему вырабатывать элементы этого множества в определенном порядке. При этом способ генерирования элементов имеет свойства, которые обычно не определяются поставленной задачей. Например, порядок, в котором вырабатываются элементы, может зависеть от "стоимости" выработки элемента и т. д. Способ проверки также требует определенных затрат времени, связанных с его применением. Задачу можно решить, если эти затраты не очень велики с точки зрения времени

и объема вычислений, требуемых для ее решения» (Аллен Ньюэлл, лауреат премии Тьюринга, член академии наук США).

«Требование, чтобы машина имела дело с неразрешимыми системами, накладывает фундаментальное ограничение на ее способ действия. Обычная методика решения задач на цифровых машинах, т. е. поиск подходящего алгоритма, теперь уже неприемлема по той простой причине, что такого алгоритма, вообще говоря, не существует. Было показано, что даже для такого простого раздела логики, как исчисление высказываний, требуется слишком много времени для нахождения доказательства алгоритмическим методом, так как при этом приходится проводить исчерпывающий поиск среди аксиом и ранее доказанных теорем, в сочетании с полным развитием "доказывающей последовательности" путем систематического применения правил преобразования до достижения требуемого доказательства. О решении таким способом задач из более интересных разделов логики заведомо не может быть и речи. Поэтому остается единственный путь – создать машину, действующую на основе эвристических методов, которые обычно используются человеком в подобных случаях» (Герберт Гелентер, инженер корпорации IBM).

«Совершенно очевидно, что алгоритмические методы, использующие прежде всего выгоды, даваемые огромной скоростью работы вычислительных машин, сами по себе не приведут к созданию разумных машин или хотя бы машин, способных заменить обычного клерка при выполнении довольно простых операций. Дело в том, что человек даже при выполнении "неинтеллектуальной работы" делает нечто большее, чем простой подсчет частот появления или поиск ключевых слов. Человек обладает интеллектуальными характеристиками, которые в общем можно выразить словами: "он понимает смысл того, что слышит или читает".

Узнавать больше, чем нам сообщили, – весьма характерное человеческое свойство, обнаруживающееся в большинстве поведений того типа, который мы называем разумным. Мы утверждаем, что эта особенность необходима и машине, предназначенной для решения действительно сложных задач по поиску информации, языковому переводу и решению проблем. Более того, нам необходимо найти эффективные способы для хранения "следствий", если мы хотим создать разумные машины с конечным объемом памяти, т. е. если мы хотим иметь действительно разумные машины» (Р. Линдсей, профессор Стэндфорского университета).

«В начале нашего века основным направлением психологии был ассоцианионизм. Это была атомистическая доктрина, которая постулировала теорию на основе небольшого числа строго определенных элементов – ощущений или представлений, которые связываются или ассоциируются друг с другом в неизменном виде. Это была механистическая теория, объяснявшая возникновение новых ассоциаций простыми и неизменными законами совпадения событий во времени и пространстве. Таковы были ее основные предположения. Согласно этой теории, поведение формируется в результате потока ассоциаций: каждая ассоциация создавала базу для других ассоциаций и достигала новых связей с "ощущениями", поступающими из внешнего мира.

В первом десятилетии нашего века (имеется в виду XX век) как реакция на эту теорию возникли работы Вюрцбургской школы. Отвергая идею всецело самодовлеющего потока ассоциаций, эта школа ввела понятие "задачи" или "цели" как необходимого фактора описания процесса мышления. Стоящая перед человеком "задача" даст направление его мышлению. Важным нововведением Вюрцбургской школы было систематическое использование самонаблюдения для выяснения процесса мышления и содержания сознания. В результате произошло некоторое слияние механизма с феноменолизмом, что, в свою очередь, привело к созданию двух совершенно противоположных направлений: бихевиоризма и гештальтпсихологии.

Бихевиористы настаивают на том, что самонаблюдение является очень ненадежной субъективной процедурой, бесплодность которой полностью обнаружилась во время дискуссий по поводу "внеобразного" абстрактного мышления. Бихевиористы по-новому сформулировали задачу психологии, как задачу выяснения реакции организма на внешние стимулы при объективном измерении как реакции, так и стимула. Однако бихевиоризм принял также и на деле подкрепил механистическую гипотезу о том, что связи между стимулом и реакцией создаются и поддерживаются только как простые детерминированные функции внешней среды.

Гештальтпсихологи, с другой стороны, встали на прямо противоположную точку зрения: они отвергли механистическую основу доктрины ассоцианионистов, но признали ценность феноменологического наблюдения. Во многом они придерживались взглядов Вюрцбургской школы, согласно которой мышление есть нечто большее, чем просто ассоциация, – мышление имеет направленность, которая определяется "задачей" или "установкой" самого субъекта. Гештальтпсихология

разработала эту доктрину совершенно новыми путями – в терминах "холистких" принципов целостной организации.

Современные психологи находятся в состоянии относительно устойчивого равновесия между полюсами бихевиоризма и гештальтпсихологии. Все мы восприняли главные уроки обеих школ: мы скептически относимся к субъективному в наших экспериментах и соглашаемся с тем, что все вводимые понятия в конечном счете должны быть доступны экспериментальному измерению бихевиористкими методами. Но мы признаем также, что человек представляет собой чрезвычайно сложно организованную систему и что простые схемы современных бихевиористов едва ли полностью ее отражают» (А. Ньюэлл, Г. Саймон).

«Что такое понятие? Обычное использование данного слова не всегда четко. Выражения вида "понятие силы", "понятие возмездия" и "понятие собаки" являются достаточно расплывчатыми. Черч предложил определение, которое было фактически принято психологами, работающими в области экспериментов по "обучению понятиям". Идея Черча заключается в том, что элементам некоторого множества объектов может быть присвоен любой символ (или наименование). Для любого произвольного объекта существует правило, касающееся описания этого объекта, с помощью которого можно решить, принадлежит ли данный объект к тому множеству объектов, для которого используется данное наименование. Правило решения в этом случае и есть "понятие" наименования, а множество объектов образует содержание этого наименования» (Эндрю Хант, программист).

«Любая проблема едва ли может вызвать у нас интерес, если мы не имеем о ней никакой информации. Обычно у нас есть какая-то база, пусть даже шаткая, для выяснения, достигнуто ли улучшение; некоторые опыты мы должны оценивать как более успешные, чем другие. Предположим, например, что мы имеем компаратор, который отбирает лучший результат из любой пары результатов пробных действий. Однако компаратор один не может сделать задачу "хорошо определенной", пока не определена цель. Но если найденная компаратором связь между попытками транзитивна (т. е. из того, что А влияет на В и В влияет на С, следует, что А влияет на С), то мы, по крайней мере, можем установить прогресс и дать машине задание за определенное время сделать лучшее, на что она способна.

Весьма важно, однако, отметить, что сам по себе компаратор, как бы он ни был совершенен, не может дать какого-либо улучшения, по

сравнению с исчерпывающим поиском. Конечно, компаратор дает нам информацию о частичном успехе, но нам необходим еще какой-то метод использования информации, который позволил бы направить поиск в перспективном для решения направлении, т. е. выбрать новые точки для пробных действий, которые в каком-то смысле были бы похожими или подобными или лежали бы в том же направлении с точками, давшими наилучшие предшествующие результаты. Для этого нам необходима какая-то дополнительная структура пространства поиска. Эта структура не должна иметь большого сходства с обычным пространственным понятием направления или расстояния, но она должна объединять точки, которые связаны эвристически...

Наиболее лобовой метод планирования, вероятно, состоит в использовании упрощенной модели проблемной ситуации. Предположим, что для данной проблемы существует некая другая проблема, существенно того же характера, но с меньшим числом деталей и меньшей сложности. Тогда мы можем сначала приступить к решению более простой проблемы. Предположим также, что при этом используется система методов, которые также проще, но до некоторой степени соответствуют оригиналу. Решение этой более простой проблемы может быть использовано как план для решения более сложной задачи...

Другой способ планирования – семантическая модель, противоположная модели гомоморфизма. В этом случае мы можем интерпретировать решаемую проблему с помощью другой, не обязательно более простой, но более знакомой, для которой уже имеются известные, достаточно мощные методы. Так, в связи с планом доказательства теоремы нам захочется знать, будут ли предложенные леммы, или "островки", в доказательстве действительно справедливы; если нет, план будет наверняка ошибочен. Часто одного взгляда на интерпретацию достаточно, чтобы сказать, будет ли данное предложение справедливым...

Я уверен, что рано или поздно мы сможем составить программу, обладающую большой способностью к решению задач, благодаря сложным комбинациям эвристических механизмов – многократной оптимизации, методов распознавания, алгебры планирования, процессов рекурсивного управления и т. п. Ни в одном из них мы не найдем местонахождения интеллекта. Что же такое в действительности интеллект? С моей точки зрения, это скорее вопрос эстетики или самолюбия, чем науки и техники! Для меня интеллект означает едва ли больше, чем комплекс активности, который мы уважаем, но не понимаем. Точно так же дело обстоит с "глубиной" в математике.

Как только понято доказательство некоторой теоремы, ее содержание кажется тривиальным. При этом может сохраняться чувство восхищения перед тем, каким образом это доказательство было открыто...

Если некая машина может ответить на вопрос о каком-то гипотетическом эксперименте, не проводя этого эксперимента, то ответ должен быть получен от некоей подмашины внутри этой машины. Выход такой подмашины (выражающий правильный ответ), так же как и ее вход (выражающий правильный вопрос), должен быть готовым описанием соответствующих внешних событий или классов событий. С точки зрения этой пары кодирующего и декодирующего каналов, внутренняя подмашина действует как среда и поэтому имеет характер модели. Проблема индуктивного вывода истинных суждений тогда может рассматриваться как проблема построения такой модели.

В той мере, в какой действия машины влияют на среду, эта внутренняя модель мира будет нуждаться в сведениях о самой себе. Если кто-то спрашивает машины, почему они решили сделать так-то и так-то (или машина спрашивает об этом сама себя), то любой ответ должен прийти от этой внутренней модели. Таким образом, доказательность интроспекции сама по необходимости должна в конечном итоге базироваться на процессах, используемых при построении кем-то образа "самого себя". Спекулятивные рассуждения о форме такой модели приводят к любопытному предсказанию, что разумные машины неохотно будут верить в то, что они только машины. Доводы следующие: наша модель "самого себя" имеет существенно двойственный характер. Имеется часть, касающаяся физической или механической среды, – с поведением неодушевленных объектов, и часть, связанная с социальной и психологической стороной. Мы вынуждены рассматривать обе эти части порознь именно по той причине, что до сих пор не имеем удовлетворительной физической теории умственной деятельности. Мы не можем отказаться от такого разделения, даже если бы желали этого, до тех пор, пока мы не создадим взамен объединенную модель. Если мы теперь спросим такую машину, что она за существо, она не сможет ответить прямо на вопрос, она должна исследовать свою модель, и она должна ответить, что она, по-видимому, является чем-то двойственным, имеющим две части – "душу" и "тело". Таким образом, даже робот, если его не снабдить удовлетворительной теорией искусственного мышления, сохранил бы дуалистический взгляд в этом вопросе» (Минский Марвин Ли, лауреат премии Тьюринга, основатель Лаборатории искусственного интеллекта в Массачусетском технологическом институте).

«Моя точка зрения состоит в том, что искусственный интеллект представляет собой инженерную дисциплину, поскольку его первоначальной целью является создание конструкций. Поэтому в поисках общей теории искусственного интеллекта смысла не больше, чем в поисках, скажем, теории гражданского строительства. Вместо единой общей теории есть ряд дисциплин, которые сюда относятся и которые должны изучаться теми, кто выбирает искусственный интеллект сферой своей деятельности. К таким дисциплинам относятся математическая логика, структурная лингвистика, теория вычислений, теория информационных структур, теория управления, статистическая теория классификации, теория графов и теория эвристического поиска» (Нильс Нильсен, ведущий сотрудник Группы искусственного интеллекта Стэндфорского университета).

«Искусственный интеллект уже сегодня превосходит человеческий во многих областях. Так, на протяжении многих лет разные виды искусственного интеллекта побеждают чемпионов всевозможных игровых турниров, будь то шахматы или покер. Такие достижения могут и не казаться особенно впечатляющими, но лишь потому, что наши требования к удивительному быстро адаптируются к прогрессу.

Для нас важно создать искусственный интеллект, у которого хватит ума учиться на своих ошибках. Он будет способен бесконечно совершенствовать себя. Первая версия сможет создать вторую, которая будет лучше, а вторая, будучи умнее оригинала, создаст еще более продвинутую третью и так далее. В определенных условиях такой процесс самосовершенствования может повторяться до тех пор, пока не будет достигнут интеллектуальный взрыв – момент, когда интеллектуальный уровень системы подскочит за короткое время с относительно скромного уровня до уровня суперинтеллекта.

Искусственный интеллект может быть менее человечен, чем пришелец. Нет ничего удивительного, что любого разумного пришельца могут побуждать к действию такие вещи, как голод, температура, травмы, болезни, угроза жизни или желание завести потомство. Искусственный интеллект, по сути, ничего из перечисленного интересовать не будет. Можно представить существование искусственного интеллекта, чьей единственной конечной целью будет пересчитать все песчинки на острове Боракай или найти десятичное представление числа π» (Ник Бостром, философ, профессор Оксфордского университета).

«Во время моего диссертационного исследования в 80-х годах я начал думать о рациональном принятии решений и о проблеме практи-

ческой невозможности оного. Если бы вы были рациональны, вы бы полагали: вот мое текущее состояние, вот действия, которые я могу совершить, потом могу сделать вот это, потом это; какой путь гарантированно приведет меня к цели? Определение рационального поведения требует оптимизации всего будущего Вселенной. Это невозможно с точки зрения вычислений. Нет никакого смысла пытаться впихнуть в ИИ невозможное, поэтому я задумался о другом: как на самом деле мы принимаем решения?

Один из трюков – это думать в краткосрочной перспективе, а затем предполагать, каким может выглядеть дальнейшее будущее. Шахматные программы, например, если бы были рациональными, то разыгрывали исключительно ходы, гарантирующие мат, но они же не делают этого. Вместо этого они просчитывают десятки ходов вперед и делают предположения о том, насколько полезны эти грядущие состояния, а затем принимают решения, которые ведут к одному из наилучших состояний.

Другое, о чем важно не забывать, – это проблема решений на множественных уровнях абстракции, "иерархическое принятие решения". Человек совершает порядка 20 триллионов физических действий за свою жизнь. Чтобы прийти на эту конференцию и рассказать что-то, понадобилось совершить 1,3 миллиарда действий или около того. Если бы вы были рациональными, вы бы пытались просчитать все 1,3 миллиарда шагов – что, по сути, совершенно невозможно. Люди делают это благодаря огромному багажу абстрактных действий высокого уровня. Вы не думаете, "когда я зайду домой, я поставлю левую ногу или правую на коврик, а затем смогу то-то и то-то". Вы думаете: "я пойду в магазин и куплю книгу. Потом вызову такси". И все. Вы вообще не задумываетесь особо, пока не требуется уточнение сложных деталей. Так и живем, в принципе. Будущее размыто, очень много деталей близко к нам во времени, но в основном представлено большими кусками вроде "получить степень", "завести детей".

Вот один из недостающих кусков головоломки: откуда берутся все эти действия на высоком уровне? Мы не думаем, что программы вроде сети DQN имеют абстрактные представления о действиях. Есть много игр, которые DQN просто не понимает, и эти игры сложные, требующие многих и многих шагов, просчитанных наперед, в виде примитивных представлений о действиях – что-то типа, когда человек думает "что мне нужно, чтобы открыть дверь?", а открытие двери включает извлечение ключа и так далее. Если у машины нет такого представления – "открыть дверь", она не сможет ничего сделать.

Но если эта проблема будет решена – а это вполне возможно, – тогда мы увидим очередной существенный рост возможностей машин. Есть две или три проблемы, и если решить их все, тогда мне непонятно, останутся ли еще крупные препятствия до ИИ человеческого уровня» (Стюарт Рассел, профессор Калифорнийского университета).

«Искусственные нейронные сети индуцированы биологией, так как они состоят из элементов, функциональные возможности которых аналогичны большинству элементарных функций биологического нейрона. Эти элементы затем организуются по способу, который может соответствовать (или не соответствовать) анатомии мозга. Несмотря на такое поверхностное сходство, искусственные нейронные сети демонстрируют удивительное число свойств, присущих мозгу. Например, они обучаются на основе опыта, обобщают предыдущие прецеденты на новые случаи и извлекают существенные свойства из поступающей информации, содержащей излишние данные.

Несмотря на такое функциональное сходство, даже самый оптимистичный их защитник не предположит, что в скором будущем искусственные нейронные сети будут дублировать функции человеческого мозга. Реальный "интеллект", демонстрируемый самыми сложными нейронными сетями, находится ниже уровня дождевого червя, и энтузиазм должен быть умерен в соответствии с современными реалиями. Однако равным образом было бы неверным игнорировать удивительное сходство в функционировании некоторых нейронных сетей с человеческим мозгом. Эти возможности, как бы они не были ограничены сегодня, наводят на мысль, что глубокое проникновение в человеческий интеллект, а также множество революционных приложений могут быть не за горами» (Ф. Уоссермен, американский ученый, специалист в области нейронных сетей).

«Одна из наиболее заманчивых перспектив, открывающаяся, таким образом, перед нами, – это рациональное руководство человеческими делами, и в частности делами, затрагивающими общественные коллективы, которые, по-видимому, представляют собой некоторую статистическую закономерность, как, например, руководство общественным явлением создания мнения. Нельзя ли представить себе машину, накапливающую тот или иной тип информации, как, например, информацию о производстве и рынке, и затем в качестве функции обычной человеческой психологии и количеств, которые можно измерить в определенном случае, определяющей наиболее вероятное развитие ситуации? Разве нельзя представить себе государствен-

ный аппарат, охватывающий все системы политических решений, либо при режиме многих государств, существующих на нашей планете, либо при явно более простом режиме общечеловеческого правительства всей планеты? В настоящее время ничто не мешает нам предположить это. Мы можем мечтать о том времени, когда machine of gouverner сможет заменить – на пользу или во вред – современную очевидную недостаточность мозга, когда последний имеет дело с обычной машиной политики.

Во всяком случае, человеческая действительность не допускает такого острого и очевидного решения, какое допускают вычисления цифровых данных. Она позволяет только определить сомнительный характер своих ценностей. Машина, обрабатывающая данные об этих процессах и поставленных ими проблемах, должна поэтому обладать вероятностной, а не детерминированной мыслью, как это представлено, например, в современных вычислительных машинах. Это делает ее задачу более сложной, однако не невозможной. Упреждающая машина, которая определяет действенность огня зенитной артиллерии, является примером этому. Теоретически упреждение времени не является невозможным, также не является невозможным определение наиболее благоприятного решения, по крайней мере, в известных пределах. Возможность создания игровых машин, как, например, играющих в шахматы, очевидно, указывает на возможность такого упреждения. Ибо человеческие процессы, составляющие объект правительственной деятельности, могут быть уподоблены играм в том смысле, в котором фон Нейман исследовал их математически. Даже если бы эти игры имели неполный комплекс правил, то существуют другие игры с очень большим числом игроков, где данные чрезвычайно сложны. Machines of gouverner будут характеризовать государство как игрока, наилучшим образом» (Н. Винер, американский математик и философ, основатель кибернетики, профессор Гарвардского, Корнельского, Колумбийского, Брауновского, Геттингенского университетов, лауреат национальной научной медали США, высшей научной награды в Соединенных Штатах Америки).

«Я еще не говорил об искусственном интеллекте. Что ж, этот предмет связан с другими политическими осложнениями, поскольку он стал частью заокеанской полемики: этот вопрос никогда не поднимался в Европе. В течение первых двух послевоенных десятилетий этому существовало простое финансовое объяснение. Искусственный интеллект был дорог, а Европа была бедна; к тому же искусственный ин-

теллект финансировался почти исключительно министерством обороны, которое направило свои усилия на субсидирование – не могу сказать "поддержку" – американских исследований. Но одна лишь финансовая сторона не объясняет всего, поскольку когда Европа стала достаточно богатой, чтобы финансировать собственные исследования в области искусственного интеллекта, об этом по-прежнему не велось и речи. На самом деле подобная участь постигла и другие отрасли науки программирования.

Мое заключение таково: это лишь один из аспектов намного более существенных культурных различий. Европейский разум поддерживает большее различие между Человеком и Машиной и ждет меньшего от обоих. Он менее склонен описывать человеческую психику в механистических терминах; он также менее склонен описывать бездушные машины антропоморфной терминологией; следовательно, он считает вопрос, может ли машина мыслить, столь же уместным, сколь вопрос, может ли субмарина плавать. Это общество, очевидно, менее тяготеет к разным техническим штучкам, отчасти потому, что не ждет от них слишком многого, тем более спасения. Наоборот, попытка имитировать человеческий разум вызывает у них лишь комментарий: "А может, попробуете скопировать что-нибудь получше?"

Обычно мне не нужно говорить об искусственном интеллекте, поскольку это всего лишь специфическая область потенциального применения машин и поскольку она находится за пределами собственно информатики. Впрочем, я вынужден делать это, как только высказывается мнение, что с применением техники искусственного интеллекта машины справятся с проблемами программного обеспечения, которые нам самим не по зубам. Нашей первой реакцией на Проект пятого поколения был вздох облегчения в таком духе: "Что ж, если японская промышленность пытается вложиться в искусственный интеллект, уж он-то позаботится о японской конкурентоспособности". Примерно через неделю пришло грустное понимание, что западному миру, похоже, не хватит силы духа удержаться и не примкнуть к этому поветрию» (Эдсгер Дейкстра, нидерландский ученый, лауреат премии Тьюринга).

«Проблема состоит в том, что пока мы не можем в целом определить, какие вычислительные процедуры мы хотим называть интеллектуальными. Мы понимаем некоторые механизмы интеллекта и не понимаем остальных. Поэтому под интеллектом в пределах этой науки понимается только вычислительная составляющая способности

достигать целей в мире. Эта неопределенность привела к тому, что задача конструирования искусственного интеллекта в своей самой общей постановке уже не имеет большого смысла. Говоря об интеллекте, равном человеческому, мы не вполне понимаем, что мы этим желаем сказать, что вообще это означает. Поэтому пока нет удовлетворительного определения, пока не задано четкое условие задачи, можно говорить об искусственном интеллекте только как о технологии моделирования деятельности, за которой мы интуитивно признаем природу интеллектуального.

Надо полагать, что пройдет довольно много времени, и наша человеческая наука пройдет большой путь, прежде чем ситуация изменится, и изменится наше понимание интеллекта вообще и искусственного в частности. Но, впрочем, нет никаких жестких оснований утверждать, что мы вообще когда-либо выйдем за пределы чисто инженерных построений, вполне возможно, что и человеческий интеллект – это тоже не более чем чисто инженерная конструкция, может быть, именно такое понимание является фундаментальным» (Джон Маккарти, лауреат премии Тьюринга, разработчик языка Лисп, профессор Стэндфорского университета).

«В начальный период развития ИИ идея применения механизмов логического вывода в аксиоматических (или квазиаксиоматических, использующих в качестве аксиом определенные законы данной предметной области) системах занимала доминирующее положение. Предполагалось, что все или почти все задачи, претендующие на интеллектуальность, можно решать путем построения некоторого вывода. Такая парадигма породила многочисленные работы в области автоматического доказательства теорем, разработки языков представления знаний логического типа, в частности хорошо известного языка Пролог. Значительные усилия были затрачены на создание методов вывода в исчислении предикатов, которое различным образом модифицировалось, чтобы адаптировать его для нужд искусственного интеллекта.

Классический подход в ИИ, реализующийся под явным давлением логических моделей в представлении знаний, породил экспертные системы, основанные на продукционных правилах, теорию реляционных баз данных, теорию решателей и планировщиков. Несомненным преимуществом, связанным с увлечением логическим выводом, было привлечение в сферу исследований области ИИ логиков, принесших в эту молодую науку свои представления о строгости и точности постановок задач и формулировок результатов.

Но уже к середине 70-х годов постепенно выясняется, что классических логических моделей и схем вывода явно не хватает для того, чтобы строить достаточно богатые и практически значимые интеллектуальные системы. Искусственный интеллект явно вырос из "логических штанишек". Принципы, опирающиеся на классическое понимание формальной системы дедуктивного вывода, стали слишком узкими для решения задач ИИ. Возникло нечто вроде кризиса в физике, ярко проявившегося в начале XX века. В чем же состояла основная проблема?

Логический подход в его классической форме требовал для каждой предметной области, для которой применялись методы ИИ, наличия полного перечня исходных положений, которые можно было бы считать аксиомами этой предметной области. Их существование (сюда, естественно, включаются и априорно задаваемые правила вывода) обеспечивало замкнутость используемых моделей, позволяло ставить и решать круг проблем, связанных с полнотой, результативностью и непротиворечивостью используемых моделей и процедур.

Однако различные приложения, к которым стремился искусственный интеллект, оправдывая свою практическую значимость, в подавляющем большинстве случаев не давали возможностей построения аксиоматических систем. Знания о предметных областях, как правило, были неполными, неточными и лишь правдоподобными, что приводило к эффектам немонотонности процессов получения результатов, возникновению фальсификаторов ранее полученных утверждений, быстрому снижению достоверности утверждений, получаемых в результате последовательного (даже при так называемых параллельных модификациях) процесса логического вывода.

Так возникла проблема замены формальной системы с присущими ей процедурами дедуктивного вывода иной, столь же мощной моделью, где отражались бы основные особенности поиска решения в плохо определенных предметных областях, которые описываются как открытые системы с обновляемыми знаниями об их строении и функционировании.

С конца 70-х годов XX века старая парадигма, опирающаяся на идею строгого логического вывода, начинает постепенно сменяться новой парадигмой, провозглашающей, что основной операцией при поиске решения должна быть правдоподобная аргументация. Работа с аргументами "за" и "против", снабженными соответствующими весами, приводит к аддитивным процедурам с этими весами (в противовес мультипликативным процедурам вычисления обобщенных весов

при правдоподобном выводе). Это обстоятельство оказалось решающим для перехода к аргументации в интеллектуальных системах.

Однако, в отличие от завершенной структуры логического вывода, до сих пор не существует столь же стройной, научно разработанной теории правдоподобной аргументации. Эффекты, связанные с появлением парадоксов при немонотонных рассуждениях, показывают, что переход к более богатой, по сравнению с моделью логического вывода, модели правдоподобной аргументации неизбежно приводит к большому количеству новых проблем, связанных с обоснованием подобной модели и изучением ее особенностей. Предполагается, что в ближайшие десятилетия усилия многих специалистов сосредоточатся именно в этой области исследований» (Дмитрий Поспелов, советский ученый, доктор технических наук, заместитель председателя Государственного научного совета по проблематике искусственного интеллекта).

«У понятия "искусственный интеллект" много объяснений. Одно из старейших дано Алленом Ньюэллом и Гербертом Саймоном, двумя пионерами исследований в этой области, которые в 1975 году написали: "Задача интеллекта заключается в том, чтобы не допустить экспоненциального взрыва при поиске решения проблемы". Они имели в виду, что существует огромное количество в большинстве своем интересных проблем, которые заставляют нас изучить экспоненциально огромное количество их потенциальных решений, чтобы найти то, которое подойдет, в случае если применить к ним метод простого перебора. Лучшим способом избежать этого "взрыва" возможных вариантов решения является интеллектуальный выбор вашей стратегии поиска. С этой точки зрения, искусственный интеллект – это наличие такой стратегии у машины, чаще всего у компьютера или у робота, который контролируется компьютером...

Есть миф о том, что роботы захватят мир и что в скором времени искусственный интеллект будет представлять серьезную опасность для нашего существования. В теории это может произойти – если бы у нас появились системы искусственного интеллекта, которые были бы настолько "умны", что могли бы самостоятельно себя перепрограммировать, чтобы еще больше развить свой интеллект, постепенно превзойти наш интеллект и нашу способность их контролировать. Так что это не совсем миф (в конце концов, Джон фон Нейман впервые использовал термин "сингулярность" в контексте возможности таких событий), но при нынешнем уровне развития искусственного

интеллекта это невозможно. Нам предстоит пройти долгий путь, прежде чем можно будет всерьез рассматривать эту возможность...

Если мы можем о чем-то мечтать, то мы можем претворить эти мечты в жизнь (конечно, если позволяют законы физики), так что я думаю, что большинство представлений о будущем искусственного интеллекта, представленных в научной фантастике, может стать реальностью. Но мы еще далеки от их претворения в жизнь. Сейчас мы находимся на уровне ньютоновской механики в области искусственного интеллекта, однако действительно сложным искусственным когнитивным системам понадобится эквивалент теории относительности или квантовой механики. Пока мы не откроем этих теорий, мы все равно сможем создавать системы искусственного интеллекта для практических целей, вот только, возможно, мы не заметим, что это настоящий искусственный интеллект» (Дэвид Вернон, профессор информатики в исследовательском центре Университета Сковдэ, координатор Европейского сообщества развития систем искусственного интеллекта).

Литература

1. *Бостром Н.* Искусственный интеллект. Этапы. Угрозы, Стратегии. М.: Манн, Иванов и Фербер, 2014.
2. Вычислительные машины и мышление / под ред. Э. Фейгенбаума и Дж. Фельдмана. М.: Мир, 1967.
3. *Гренандер У.* Лекции по теории образов. М.: Мир, 1979.
4. *Нильсен Н.* Искусственный интеллект. М.: Мир, 1973.
5. *Нильсен Н.* Кибернетика. М.: Издательство иностранной литературы, 1968.
6. *Поспелов Д.* Моделирование рассуждений. Опыт анализа мыслительных актов. М.: Радио и связь, 1989.
7. *Поспелов Д.* Фантазия или наука: на пути к искусственному интеллекту. М.: Наука, 1982.
8. *Рассел С., Норвиг П.* Искусственный интеллект. Современный подход. М.: Вильямс, 2006.
9. *Уоссермен Ф.* Нейрокомпьютерная техника. Теория и практика. М.: Мир, 1992.
10. *Финн В.* Автоматическое порождение гипотез в интеллектуальных системах. М.: Либроком, 2009.
11. *Хант Э.* Искусственный интеллект. М.: Мир, 1978.

Книги издательства «ДМК Пресс» можно заказать
в торгово-издательском холдинге «Планета Альянс»
наложенным платежом,
выслав открытку или письмо по почтовому адресу:
115487, г. Москва, 2-й Нагатинский пр-д, д. 6А.
При оформлении заказа следует указать адрес (полностью),
по которому должны быть высланы книги;
фамилию, имя и отчество получателя.
Желательно также указать свой телефон и электронный адрес.
Эти книги вы можете заказать и в интернет-магазине: **www.alians-kniga.ru**.
Оптовые закупки: тел. **(499) 782-38-89**.
Электронный адрес: **books@alians-kniga.ru**.

Потопахин Виталий Валерьевич

Романтика искусственного интеллекта

Главный редактор *Мовчан Д. А.*
dmkpress@gmail.com
Корректор *Синяева Г. И.*
Верстка *Чаннова А. А.*
Дизайн обложки *Мовчан А. Г.*

Формат 60×90 1/16.
Гарнитура «Петербург». Печать офсетная.
Усл. печ. л. 10,625. Тираж 200 экз.

Веб-сайт издательства: www.дмк.рф

Романтика искусственного интеллекта

Эта книга о том, чем занимаются специалисты по искусственному интеллекту. О том, в решении каких задач умные машины уже заменили человека, и какие интеллектуальные технологии могут появится в обозримом будущем. О том, может ли машина стать равноценным партнером человека или даже превзойти его? Насколько реальна возможность бунта машин, так любимого писателями-фантастами? А может быть, искусственный интеллект – это просто область технического моделирования поведения, которое мы считаем разумным? И как понять, что умные машины уже живут рядом с нами?

Издание предназначено для широкого круга читателей, интересующихся вопросами искусственного интеллекта.

Сайт автора – www.lotos-khv.ru

Книга о интеллекте, проблемах его разработки и критериях его распознания.

Интернет-магазин:
www.dmkpress.com
Книга – почтой:
e-mail: orders@alians-kniga.ru
Оптовая продажа:
«Альянс-книга»
Тел./факс: (499) 782-3889
e-mail: books@alians-kniga.ru

ДМК ПРЕСС
ИЗДАТЕЛЬСТВО
www.дмк.рф

ISBN 978-5-97060-476-2