

Роман Душкин

Искусственный интеллект



Роман Душкин

Искусственный интеллект

«ДМК Пресс»

2019

УДК 004.89
ББК 32.813

Душкин Р. В.

Искусственный интеллект / Р. В. Душкин — «ДМК Пресс», 2019

ISBN 978-5-97060-787-9

Перед вами книга по искусственному интеллекту от известного технологического евангелиста Романа Душкина, директора по науке и технологиям Агентства Искусственного Интеллекта*. В этой книге вы найдёте исчерпывающее описание современного состояния технологий искусственного интеллекта и сфер жизни, где их можно применять. Более того, автор сделал упор на гуманитарной составляющей исследований в области искусственного интеллекта, что выгодно отличает эту книгу от множества других изданий по теме. Также здесь развенчиваются многочисленные мифы об искусственном интеллекте и описывается авторское видение будущего. Издание будет интересно всем, кто хочет быстро погрузиться в горячую тему искусственного интеллекта, получить базовую терминологию и освоить основные методы.

УДК 004.89

ББК 32.813

ISBN 978-5-97060-787-9

© Душкин Р. В., 2019

© ДМК Пресс, 2019

Содержание

Предисловие	6
Предисловие от издательства	9
Отзывы и пожелания	9
Список опечаток	10
Нарушение авторских прав	11
Глава 1	12
Глава 2	36
Раздел 2.1. Символьные вычисления	40
Раздел 2.2. Искусственные нейронные сети	43
Конец ознакомительного фрагмента.	57

Роман Душкин

Искусственный интеллект

© Душкин Р. В., 2019 ISBN 978-5-97060-787-9

© Издание, ДМК Пресс, 2019

* * *



***Роман Душкин** – директор по науке и технологиям Агентства Искусственного Интеллекта, член Российской Ассоциации Искусственного Интеллекта. Его курсы по ИИ, машинному обучению, разговорным интерфейсам, философии сознания и другим темам опубликованы на Coursera, Udey, GeekBrains, а также преподаются в Университете Национальной Технологической Инициативы и Российской Академии Народного Хозяйства и Государственной Службы.*

Предисловие

Как всегда, начиная новую книгу, я испытываю двойственные чувства. С одной стороны, хочется написать по-настоящему полезную книгу, отражающую всё накопленное к текущему моменту знание по заявленной теме. С другой стороны, всегда опасаясь, что полученная в итоге рукопись сразу же окажется устаревшей, ведь сегодня технологии несутся в будущее со скоростью, превышающей возможности отдельного человека. И осознание этого немного удручает.

Тем не менее сегодня я сажусь за новую книгу, в которой постараюсь раскрыть неискущённому читателю тему искусственного интеллекта. Ведь не секрет, что в последние годы вокруг этого понятия поднялась невообразимая шумиха, которая доносится как из средств массовой информации, так и из деловой среды, в которой что ни день открывается новый стартап, основанный на самых современных технологиях искусственного интеллекта. Подогревает эту ситуацию и массовая культура, где тема ИИ тоже стала горячей. Появляется большое количество новых произведений, в которых человеческое общество так или иначе взаимодействует с порождённым (или самозародившимся) искусственным интеллектом.

Впрочем, в киноиндустрии и художественной литературе тема искусственного интеллекта эксплуатировалась с давних времён. Первые писатели-фантасты уже задумывались об искусственных существах, наделённых разумом и пытающихся либо помогать людям, либо столкнуть человека с пьедестала «царя природы» и «венца творения». Дальше – больше. Серия фильмов про терминаторов, затем красочно-философские фильмы «Матрица» необычайно подогрели интерес к этой теме. Художественная литература тоже не отставала. Если не углубляться, то на поверхности лежат трилогия «Киберпространство» Уильяма Гибсона и сага-тетралогия «Гиперион» Дэна Симмонса. Не побоюсь сказать, что только эти два произведения оказали на развитие понимания искусственного интеллекта у массового читателя большее влияние, чем все остальные книги вместе взятые.

Вместе с тем отношение в научной и инженерной среде к термину «искусственный интеллект» и этому направлению вообще несколько предвзято. Всё дело в так называемых «зимах искусственного интеллекта», т. е. практически полных остановках исследований из-за отсутствия финансирования и разочарования пионеров новой науки. На смену первым восторженным надеждам пришло горькое осознание того, что человек ещё очень далёк от понимания природы сознания и всех тех особенностей мозга, которые делают человека разумным существом. К тому же ещё двадцать лет назад не было достаточных вычислительных мощностей для реализации всех тех теоретических находок, которые были сделаны в научных лабораториях. Так что в среде понимающих специалистов отношение к теме скорее скептическое. Кроме этого, развитие искусственного интеллекта как междисциплинарного направления исследований уже пережило две зимы.

Причиной шумихи вокруг искусственного интеллекта, начавшейся во втором десятилетии XXI века, как видится, являются два процесса. Во-первых, вычислительные мощности и объём имеющихся в распоряжении человечества в целом вычислительных устройств достигли небывалых размеров, и этот размер имеет тенденцию увеличиваться по экспоненциальному закону. Сегодня количество смартфонов, которые можно связать в грид для распределённых вычислений (причём часто без ведома владельцев), достигло двух миллиардов, а каждый смартфон обладает мощностью, на порядки превышающей мощность тех персональных компьютеров, которые были в распоряжении учёных ещё двадцать пять лет назад.

Во-вторых, выросло, получило образование и начало усердно работать поколение людей, которые застали вторую зиму искусственного интеллекта еще младенцами. Но сегодня, получив образование, намного более серьёзное, чем предшественники, представители этого поко-

ления с удесятерёнными силами ухватились за старые надежды, пренебрегая тем скепсисом, который всё ещё имеется у представителей «старой школы». Это и хорошо, и не очень. Хорошо потому, что, не будучи зашоренными, новые специалисты могут «перепрыгнуть» барьер недоверия, выстроенный вокруг искусственного интеллекта. Не очень потому, что многие наступят на те же грабли, что и исследователи первой половины XX века.

Из этих двух тезисов и родилась идея книги. Я хотел бы дать полноценное описание современного состояния технологий искусственного интеллекта, чтобы оно стало основой для дальнейшего развития новых специалистов, которые уже смогли бы сдать в утиль старые парадигмы и пройти узкой тропинкой к созданию настоящего искусственного интеллекта, так как любая разумная раса в конечном итоге должна задуматься о природе своего разума и стремиться к созданию разума искусственного.

В книге мы обсудим все аспекты искусственного интеллекта. Начнём с истории развития этой области знаний. Поговорим о том, как всё начиналось и какие направления были открыты и потом закрыты по мере развития понимания проблемы. Мы изучим все предпосылки, послужившие формированию новых технологий. В конце первой главы мы изучим типы искусственного интеллекта и получим все необходимые знания, для того чтобы перейти к рассмотрению технологий и их современного состояния.

Во второй главе мы кратко рассмотрим три основополагающих современных метода построения искусственных интеллектуальных систем – символьные вычисления и логический вывод, искусственные нейронные сети и эволюционные алгоритмы. Фактически сегодня это три столпа искусственного интеллекта, на которых базируются прикладные направления исследований.

Далее в третьей главе книги будет представлено более или менее подробное описание технологий искусственного интеллекта и их текущее состояние. Мы начнём с двух направлений, которые развивались в рамках исследований по искусственному интеллекту, и постепенно перейдём к исследованиям, находящимся сегодня на острие науки. В разделе рассмотрим такие темы, как машинное обучение, искусственные нейронные сети, символьные вычисления, методы представления знаний, обработка естественного языка, робототехника и бионика и др.

Четвёртая глава посвящена философии искусственного интеллекта. Мы рассмотрим разные модели сознания, изучим философские подходы к изучению его природы. Это позволит понять, насколько мы ещё далеки от понимания того, что представляет собой разум человека и где он может находиться. Вместе с тем изучение философии искусственного интеллекта позволит очертить те направления движения, куда необходимо двигаться для постижения сути интеллекта и осуществления попыток его разработки в виде искусственной системы.

В пятой главе представлены описания и объяснения всевозможных мифов и опасений человечества относительно искусственного интеллекта и его взаимодействия с людьми и всем человеческим обществом в целом. В массовом сознании сформировалось некоторое количество страхов, которые сегодня могут вылиться в попытки государств или даже надгосударственных структур ограничить или как-то регламентировать исследования в этой области знаний, как это ранее было сделано с генетикой и точными медицинскими технологиями. Несомненно, какое-то регламентирование необходимо, но оно должно быть основано на здравом понимании, а не на страхах и мифах.

Шестая глава посвящена вопросам применения технологий искусственного интеллекта в различных сферах жизни – медицине, образовании, обеспечении безопасности и многих других. В разделе описаны как уже имеющиеся применения технологий, так и варианты будущих применений при развитии методов. Этот раздел будет наиболее интересен тем читателям, которые ищут новые идеи и возможности по изменению существующего порядка вещей.

Седьмая глава описывает перспективы и прогнозы технологий искусственного интеллекта. Что они дадут конкретным людям и человечеству в целом, куда приведут нас как био-

логический вид, и наконец, какие параллельные исследования получают второе дыхание. Мы рассмотрим футуристические предположения о том, что нам даст искусственный интеллект и когда это может произойти.

Наконец, тех, кто прочитал все шесть разделов до конца, в конце книги ждёт небольшой сюрприз.

В добрый путь...

Москва, 2017–2019 гг.

Душкин Р. В.

Предисловие от издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв прямо на нашем сайте **www.dmkpress.com**, зайдя на страницу книги, и оставить комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу **dmkpress@gmail.com**, при этом напишите название книги в теме письма.

Если есть тема, в которой вы квалифицированы, и вы заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу **http://dmkpress.com/authors/publish_book/** или напишите в издательство по адресу **dmkpress@gmail.com**.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы удостовериться в качестве наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в тексте или в коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от расстройств и поможете нам улучшить последующие версии этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу **dmkpress@gmail.com**, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконно выполненной копией любой нашей книги, пожалуйста, сообщите нам адрес копии или веб-сайта, чтобы мы могли применить санкции.

Пожалуйста, свяжитесь с нами по адресу электронной почты **dmkpress@gmail.com** со ссылкой на подозрительные материалы.

Мы высоко ценим любую помощь по защите наших авторов, помогающую нам предоставлять вам качественные материалы.

Глава 1

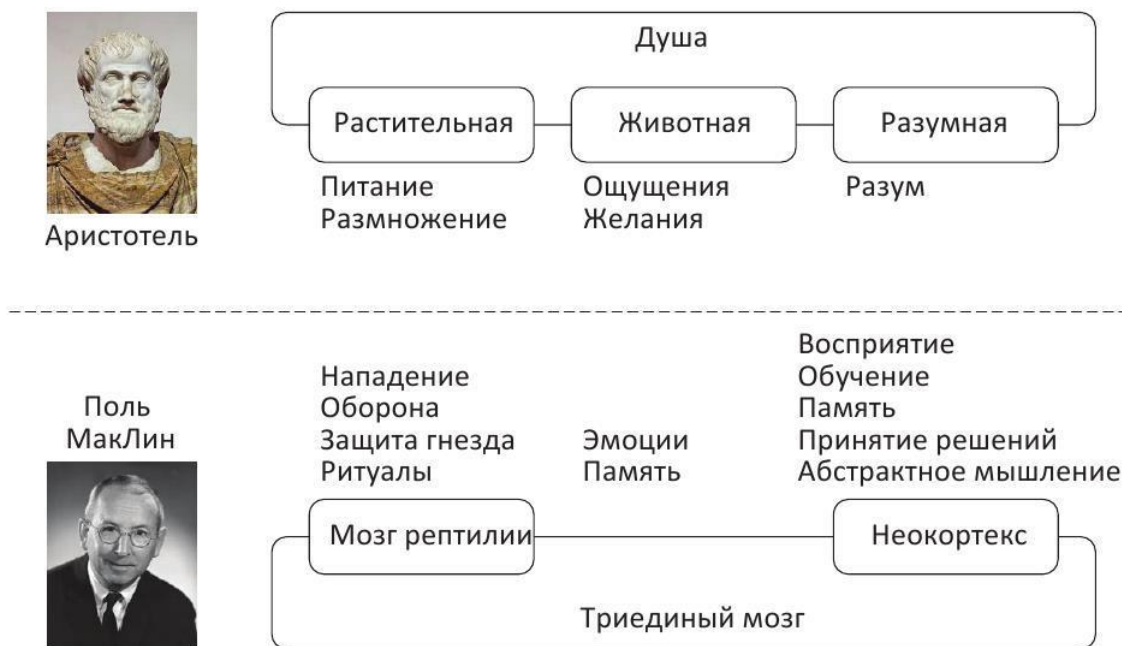
История развития

Похоже, что природа разума и сознания взволновала человека прямо сразу, как только он смог осознать свою личность и процесс мышления. Научных свидетельств этому нет, однако можно предположить, что все эти отсылки к «душе», «искре божьей», «духу» и другим подобным штукам имеют в своей основе попытки найти то самое сокровенное, чем человек отличается от «неразумных животных». Конечно же, это относится к разуму, относительно которого до сих пор невозможно сказать как он работает, где скрыт, чем обусловлен. Поскольку подступиться к научному решению этого вопроса до возникновения естественно-научного мировоззрения было невозможно, любопытным людям приходилось плодить сущности.

Вместе с тем развитие научного метода и философии науки происходило в рамках европейской цивилизации, поэтому дальнейшее изложение будет проводиться с европоцентристской точки зрения. Такая позиция не должна отпугивать читателя, поскольку современные достижения науки своей основой лежат именно в европейской колыбели, имея лишь небольшие вкрапления знаний иных цивилизаций. Несмотря на то что такие древние цивилизации, как Индия, Китай и арабский мир, принесли много интересных находок и идей, все они были переосмыслены в европейской традиции, после чего влились в единый поток научного знания.

Поскольку и европейская цивилизация, и современная наука имеют своим истоком древнегреческую философию, имеет смысл обратиться к ней. Особый интерес представляет классический период древнегреческой философии, одним из ярчайших представителей которого является Аристотель. Именно в его трудах складывается научная система мира и познания объективной реальности. Не зря вся современная наука базируется на той самой формальной логике, которую разработал и детально описал Аристотель.

Интересно то, что, согласно Аристотелю, душа человека вмещает в себя три части – растительную, животную и разумную. Первая часть души отвечает за такие функции, как питание и размножение. Вторая часть, животная, – ощущение и желание. Наконец, разумная часть души имеется только у человека, и именно в ней находится разум. Это понимание очень похоже на триединую модель мозга, предложенную в 60-х годах XX века американским нейрофизиологом Полем МакЛином. В этой модели мозг разделён на три взаимосвязанные и тесно взаимодействующие части – так называемый «мозг рептилии», лимбическую систему и неокортекс. Собственно, мозг рептилии – это самая древняя часть головного мозга, которая имеется у рептилий и птиц и отвечает за базовые функции, связанные с инстинктивным поведением различного плана: нападательным, защитным, территориальным и ритуальным. Лимбическая система отвечает за эмоциональную составляющую жизнедеятельности животного, а также за отдельные виды памяти. Наконец, неокортекс свойствен только высшим животным и отвечает за такие функции, как восприятие и самовосприятие, обучение и память, принятие решений и абстрактное мышление. Предполагается, что именно в неокортексе располагается сознание человека.



Соответствие моделей мозга Аристотеля и Поля МакЛина

Другими словами, только лишь по внешним наблюдениям и при помощи спекулятивных размышлений Аристотелю удалось предвосхитить одну из наиболее интересных моделей мозга, принятых уже в наше время. Впрочем, это действительно можно сделать, наблюдая и сопоставляя поведение различных живых организмов. Из-за того что у древних мыслителей всё-таки не было методов изучения человеческого мозга во всех его аспектах и функциональности, поискам природы разума посвящалось незначительное количество усилий. А в темные времена, наступившие в Европе, этот вопрос был совсем отставлен. Потихоньку изучение человеческого интеллекта и его сущности начало выходить на повестку дня только во времена эпохи Возрождения. Но и тогда, несмотря на становление философии гуманизма, философам и мыслителям было сложно подступить к этой теме.

Прорывом стала книга Рене Декарта «Рассуждения о методе, чтобы хорошо направлять свой разум и отыскивать истину в науках», которую он опубликовал в 1637 году. Несмотря на то что в то ужасное время гонения на свободных разумом учёных было делом обыденным, Декарт отважился на публикацию этой поистине прорывной работы (хотя если говорить откровенно, на современном языке эту работу лучше всего назвать «подрывной»). Она стала знаковой и, по мнению многих историков науки и философии, ознаменовала собой переход к современному научному познанию и к философии Нового времени. Фактически книга «Рассуждения о методе...» стала базисом всей современной эпистемологии.

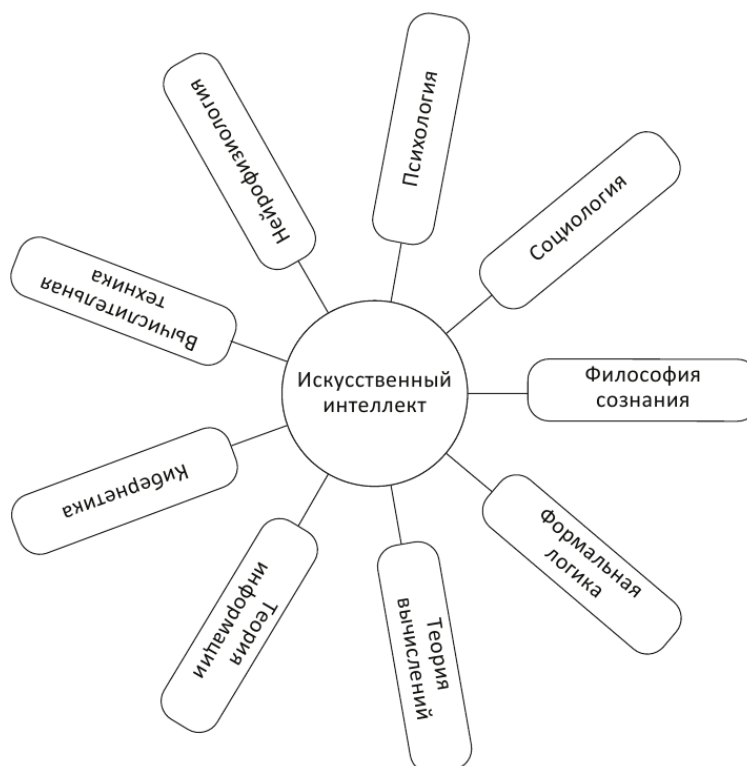
Декарт провозгласил механистический подход к пониманию природы и объективной реальности, т. е. рассмотрение объектов в окружающем мире в качестве неких механизмов, подчиняющихся физическим законам. Это позволило отвергнуть так называемый витализм, когда движущей силой живой материи объявлялась некоторая сверхъестественная сущность, не подчиняющаяся законам физики, химии, биологии. Собственно с этого момента учёные смогли переключить перспективу и начать рассматривать разум и интеллект человека в той же парадигме.

Однако дальше дело не пошло, поскольку, кроме научного метода, у учёных до второй половины XIX века просто не было достаточной инструментальной базы как с точки зрения математических формализмов для исследования феноменологии мышления, так и с точки зрения физических и химических инструментов для исследования материи мозга.

«Взрывное» развитие математики и вслед за ней прикладных наук начиная со второй половины XIX века дало стимул к развитию и различным направлениям в области изучения мышления и интеллекта. Но здесь уже сложно говорить о какой-то последовательной работе над этой темой, так как сам по себе вопрос изучения человеческого разума является междисциплинарным и про историю развития наук об искусственном интеллекте можно говорить исключительно в ключе постепенного сбора информации и разработки методов в совершенно различных областях знания, которые затем объединялись и в конечном итоге привели к выделению искусственного интеллекта в отдельное научное направление. Другими словами, искусственный интеллект, как дерево, базируется на мощной корневой системе, в которую отдельными корешками входят различные науки от философии, чистой математики и теории вычислений до нейрофизиологии и психологии.

Так что можно выделить следующие магистральные направления научной мысли, которые питают искусственный интеллект (не включая отдельные более узкие специализации):

- философия сознания;
- формальная логика;
- теория вычислений;
- теория информации;
- кибернетика;
- вычислительная техника;
- нейрофизиология;
- психология;
- социология.



Итак, исследования в рамках философии сознания начал Рене Декарт, после которого к этой теме приобщались многочисленные философы, среди которых стоит отметить Джона Локка и Дэвида Юма. Много усилий к изучению природы сознания прилагали философы немецкой школы, в частности Артур Шопенгауэр и Георг Гегель. К XX веку философия созна-

ния выделилась в отдельное направление в рамках *аналитической философии*, которое является чуть ли не единственной прикладной философской дисциплиной, имеющей большое значение для многих направлений научной мысли, в том числе и для искусственного интеллекта. Философия сознания пытается ответить на вопросы о природе сознания и его соотношения с объективной реальностью, что связано с моралью, свободой воли и этическими вопросами, которые, в свою очередь, сразу же возникают при более глубоком изучении проблем искусственного интеллекта и развития его взаимоотношений с интеллектом естественным.

Как уже было упомянуто, основы формальной логики заложил ещё в античные времена Аристотель, однако потом до начала XX века эта методология использовалась учёными без какого-либо развития. Математики пытались как-то формализовать научный метод, и даже было произведено несколько интересных попыток. Немецкие математики и логики Георг Кантор и Готлоб Фреге фактически стали отцами наивной теории множеств и теории предикатов первого порядка соответственно. Эти теории позволяли формализовать очень многое, однако страдали от важного недостатка – противоречивости. И только в 1910–1913 гг. английские математики и философы Бертран Рассел и Альфред Уайтхед опубликовали трёхтомную работу «Принципы математики», где они ввели теорию типов как инструмент для более точной формализации основ математики, в которой невозможно было сформулировать парадокс Рассела о «множестве всех множеств». Именно после этой книги развитие математики пошло семимильными шагами, в результате чего сделали свои открытия Курт Гёдель, Алонзо Чёрч, Алан Тьюринг и многие другие замечательные учёные. Так что в историческом ряду развития формальной логики стоят такие личности, как Аристотель, Г. Кантор, Г. Фреге, Б. Рассел, А. Уайтхед и К. Гёдель.

После того как в 1931 г. австрийский логик Курт Гёдель опубликовал свои работы, в которых было приведено доказательство его знаменитых теорем о неполноте, начались исследования в этом направлении. Многие из них были чисто философскими, однако две примечательные работы легли в основу всей современной вычислительной техники. Первая – лямбда-исчисление Алонзо-Чёрча, описанное в его теперь уже знаменитой статье 1936 г., в которой он показал существование неразрешимых задач. Параллельно ему Алан Тьюринг переформулировал теорему Гёделя и, пытаясь решить «проблему разрешения» Давида Гильберта, разработал формализм в виде гипотетического устройства, которое впоследствии стало носить название «машины Тьюринга».

Обобщение достижений А. Чёрча и А. Тьюринга привело к формулированию тезиса Чёрча-Тьюринга, который, являясь эвристическим утверждением, гласит, что для любой алгоритмически вычислимой функции существует вычисляющая её значения машина Тьюринга. Этот тезис постулирует эквивалентность между интуитивным понятием алгоритмической вычислимости и строго формализованными понятиями *частично рекурсивной функции* (по Чёрчу), или *функции, вычислимой на машине Тьюринга* (по Тьюрингу). Тезис невозможно строго доказать или опровергнуть ввиду того, что интуитивное понятие алгоритмической вычислимости строго не определено. Однако этот тезис в совокупности с теорией вычислений сегодня лежит в основе алгоритмического решения задач и, как следствие, имеет непосредственное применение в рамках искусственного интеллекта.

Вместе с тем в 1948 году американский математик Клод Шеннон публикует статью «Математическая теория связи», которая сегодня считается вехой в рождении теории информации. Несмотря на то что до К. Шеннона в области проблем передачи информации работали такие пионеры, как Гарри Найквист и Ральф Хартли, именно Клоду Шеннону удалось математически точно сформулировать основные положения новой науки, определить её базис и доказать основную теорему, позже названную его именем. Эта теорема определяет предел максимального сжатия данных и числовое значение информационной энтропии. В связи с дальнейшим развитием сетей передачи данных теория информации и все её приложения

стали развиваться семимильными шагами, что привело к появлению многочисленных способов помехоустойчивого кодирования информации для каналов с шумом. Всё это имеет самое непосредственное значение для развития интеллектуальных систем, поскольку вопросы передачи информации в них стоят на одном из первых мест.

Конечно, одной из центральных наук, стоящих в основе разработки технологий искусственного интеллекта, является кибернетика как базовая методология исследования сложных систем, взаимодействующих друг с другом и со средой. Кибернетика сама по себе является междисциплинарной областью исследования, базирующейся на многих отраслях науки, в том числе и уже перечисленных здесь ранее. Однако именно разработанный в её рамках научный аппарат в полной мере позволяет целенаправленно заниматься поиском и проектированием сложных адаптивных и самообучающихся систем, к которым, вне всяких сомнений, относятся системы искусственного интеллекта. Кибернетику как науку разрабатывали такие знаменитые учёные, как Уильям Росс Эшби, Карл Людвиг фон Бергаланфи, Джон фон Нейман, Стаффорд Бир, а также многочисленная когорта русских учёных, среди которых обязательно надо упомянуть Ивана Алексеевича Вышеградского (основоположника теории автоматического регулирования), Алексея Андреевича Ляпунова, Виктора Михайловича Глушкова и популяризатора науки об искусственном интеллекте Льва Тимофеевича Кузина.

Все теоретические изыскания в области теории вычислений, теории информации, кибернетики и других наук в конечном итоге приводят к развитию вычислительной техники как прикладной дисциплины, рассматривающей вопросы создания и программирования универсальных вычислительных машин. Попытки создать механическую машину для вычислений предпринимались со времён, наверное, Блёза Паскаля, и первым в этом преуспел наш соотечественник Семён Николаевич Корсаков, который в 1830-х годах создавал первые «интеллектуальные машины» на перфокартах. Хотя в те же самые годы английский математик Чарльз Бэббидж разрабатывал проект универсальной цифровой вычислительной машины, до реализации дело у него не дошло. Несмотря на всё это, основы современной вычислительной техники были заложены Джоном фон Нейманом, который разработал принципы построения архитектуры универсальных вычислительных машин. Впрочем, первый компьютер в современном понимании сделал немецкий инженер Конрад Цузе, он же разработал и первый язык программирования высокого уровня, однако из-за военно-политических особенностей мира в те времена работы Цузе оставались малоизвестными. После Второй мировой войны работы над созданием универсальных компьютеров велись во всё ускоряющемся ритме, который выдерживается до сих пор (так называемый «закон Мура»).

Однако по мере продвижения в деле создания всё более мощных вычислительных систем становилось ясно, что разработать интеллект *in silico* так просто не получится, и исследователей в этом отношении ждало такое же разочарование, как создателей различных механических кукол (автоматов), которым казалось, что ещё чуть-чуть – и их куклы обретут разум. Как не получилось с механикой, так же не получилось и с электроникой. Это способствовало вовлечению в исследования нейрофизиологов и других специалистов в части анатомии, физиологии и других аспектов функционирования нервной системы и других регуляторных систем организмов человека и животных.

И вот в 1943 г. американские нейрофизиолог Уоррен Мак-Каллок и математик Уолтер Питтс публикуют статью, которая открыла миру новую вычислительную модель, основанную на понятии искусственного нейрона. Да, эта модель была довольно упрощённой и не принимала во внимание большое количество свойств органических нейронов, однако она позволяла производить вычисления. Эта статья фактически открыла широчайшее направление исследований, которое сегодня превалирует в области искусственного интеллекта – искусственные нейронные сети. Вслед за У. МакКаллоком и У. Питтсом следует отметить таких учёных, как канадский физиолог Дональд Хебб, который описал принципы обучения искусственного ней-

рона (он предложил первый работающий алгоритм обучения искусственных нейронных сетей), а также американский нейрофизиолог Фрэнк Розенблатт, который разработал на искусственных нейронах устройство, моделирующее процесс восприятия, – перцептрон.

Но всё, как обычно, оказалось не таким простым, как казалось. Несмотря на то что исследователям удалось смоделировать один нейрон и составить из таких моделей нейронную сеть, сознания в ней так и не зародилось. С одной стороны, это было связано с тем, что на тех вычислительных мощностях, которые были доступны учёным в середине XX века, можно было смоделировать нейронную сеть, состоящую из пары сотен нейронов и нескольких слоёв. Такой объём совсем не соответствует десяткам миллиардов нейронов в головном мозге человека с сотнями тысяч связей для каждого нейрона. С другой стороны, становилось понятно, что «карта не является местностью», так что ждать самозарождения сознания в нейронной сети, даже если она будет очень сложной, слишком странно. Поэтому исследователи обратились к такой науке, как психология.

Одним из первых учёных, кто обратил внимание на этот аспект искусственного интеллекта и его отношения к человеческому разуму, был русский учёный-медик и кибернетик Николай Михайлович Амосов. В ряду его обширной библиографии есть такие знаковые работы, как «Искусственный разум», «Автоматы и разумное поведение» и «Алгоритмы разума». Несмотря на глубокую степень проработки ряда важнейших вопросов, в этих работах всё так же остаётся нераскрытым вопрос о природе сознания. И получается довольно парадоксальная ситуация. Есть практически полное понимание того, как работает нейрон на уровне клетки и субклеточных структур вплоть до биохимических реакций и метаболических путей преобразования веществ, что фактически и эмулирует вычислительные процессы в рамках одной клетки. Также есть понимание, но уже не такое целостное, того, как работают нейронные сети. А ещё есть понимание психологии человека, его поведения и высших когнитивных и интеллектуальных функций. Это понимание ещё менее глубокое, но есть множество операционных гипотез, которые позволяют описывать, объяснять и предсказывать. Но при этом нет никакого понимания того, что находится посередине. Каким образом биохимические реакции нейрона и электрохимические процессы в нейронных сетях приводят к возникновению сознания, интеллекта и разума? Ответа на этот вопрос до сих пор нет. И в итоге получается, что психология – это попытки «дизассемблировать» высшую психологическую деятельность центральной нервной системы человека, но они отвязаны от базовых химических и физических процессов.

Наконец, учёные задумались и о таком важном аспекте интеллектуальной деятельности, как «коллективный интеллект». Само человеческое общество в целом часто показывает более высокий уровень интеллектуальной деятельности, чем каждый его отдельный представитель. Ведь многие сложные научные концепции и технические объекты могут быть изобретены и разработаны исключительно в рамках объединения усилий разноплановых специалистов. Однако не только человеческое общество показывает подобный паттерн поведения. Он виден и в жизни, например, общественных насекомых, когда каждая отдельная особь, вообще не имеющая и тени разума, делает вклад в поведение своего «суперорганизма», который кажется вполне интеллектуальным. Муравьи и пчёлы – это вершина эволюции насекомых на Земле. Эти наблюдения натолкнули исследователей в области искусственного интеллекта на идеи о «роевой модели интеллекта» в рамках так называемых многоагентных систем. Это – одно из самых быстро развивающихся сегодня направлений науки и техники. Здесь сложно назвать какую-либо ключевую личность, поскольку тема возникла совсем недавно. Но эта тема очень горячая и ещё долго будет оставаться на острие научного поиска.

* * *

Итак, мы рассмотрели большую часть научных направлений, которые лежат в основе искусственного интеллекта. Это, можно сказать, его базис и научные и технологические предтечи, предпосылки. Когда же родилось само направление исследований, которое называли «искусственный интеллект»? Для ответа на этот вопрос необходимо обратиться к работам философов и практиков искусственного интеллекта Марвина Мински и Джона Маккарти. Вторым так вообще является автором самого термина «искусственный интеллект» (а кроме того, он разработал язык программирования LISP и является одним из основоположников функционального программирования). Эти учёные основали в 1959 г. Лабораторию информатики и искусственного интеллекта в рамках Массачусетского технологического института, и это была первая научная лаборатория, которая занималась данной проблемой.

Именно Джон Маккарти сформулировал основополагающие принципы искусственного интеллекта, определив то, что потом было названо «чистым подходом», или «нисходящим искусственным интеллектом», и выразилось в гипотезе Ньюэлла-Саймона о том, что осмысленные действия можно выполнять только при наличии в некоторой системе механизма символьных вычислений, а сами такие символьные вычисления являются необходимым условием наличия в этой системе интеллекта. Другими словами, подход Джона Маккарти выражался в том, что системы искусственного интеллекта должны имитировать высокоуровневые психологические процессы разумного существа, такие как логическое мышление, логический вывод, речь, творчество и т. д.

С другой стороны, его друг и коллега Марвин Мински сформулировал совершенно противоположное определение искусственного интеллекта, которое получило наименование «грязного подхода», или «восходящего искусственного интеллекта». В основе этой парадигмы лежит попытка моделирования естественных процессов, происходящих в самой природе человека. В первую очередь это, конечно же, моделирование нейросетевых процессов в разных аспектах. Наиболее широко проявившейся технологией в рамках грязного подхода стали искусственные нейронные сети, которые моделируют разные процессы человеческого разума на логическом уровне. Можно было бы попробовать смоделировать биохимический уровень, однако для этого не хватает вычислительных мощностей даже сегодня, не говоря уже про те давние времена. Другой известной технологией являются различные генетические и эволюционные методы решения задач. Но в целом этот подход не является «искусственным интеллектом» в том понимании Джона Маккарти, как он определил данный термин.

При этом необходимо понимать, что первоначально предназначением той междисциплинарной области исследований, которая получила название «искусственный интеллект», было моделирование когнитивных функций человека для их исследования на модели, чтобы понять природу интеллекта, разума и сознания человека. Другими словами, искусственный интеллект первоначально рассматривался как довольно фундаментальная область исследований, и только через какое-то время появилась задача по практическому применению наработок, которая нашла своё отражение в создании большого количества искусственных систем, решающих задачи, традиционно относившиеся к прерогативе человека.

Фактически две парадигмы, описанные выше, лежат в основе всякого подхода к разработке искусственного интеллекта. На сегодняшний день таких подходов выделяют семь:

- 1) интуитивный;
- 2) логический;
- 3) символьный;
- 4) структурный;
- 5) эволюционный;

- 6) квазибиологический;
- 7) агентный.

Надо отметить, что агентный подход к разработке искусственного интеллекта чаще всего рассматривается в рамках так называемой гибридной парадигмы, которая представляет собой смесь нисходящей и восходящей парадигм, берет из них лучшее и старается нивелировать отрицательные стороны. Гибридная парадигма и агентный подход будут рассмотрены в самом конце этой главы.

Интересно то, что в рамках искусственного интеллекта, который, как уже было отмечено, является междисциплинарным научным направлением исследований, имеется ряд задач, которые решаются методами всех или некоторых перечисленных подходов. В частности, к таким задачам традиционно относят:

- поиск информации;
- обработка естественного языка;
- представление знаний;
- машинное обучение;
- распознавание образов;
- интеллектуальный анализ данных, или «дата-майнинг»;
- обработка НЕ-факторов знания;
- принятие решений;
- робототехника;
- роевой интеллект.

Нельзя сказать, что это полный и консистентный список задач, к тому же некоторые из представленных задач в какой-то части пересекаются. Тем не менее это хороший список, являющийся анкером, от которого можно отталкиваться при изучении подходов и методов искусственного интеллекта. Другие исследователи могут предлагать иной список и классификацию задач, но далее в этой книге будет рассматриваться решение именно этих задач разными методами искусственного интеллекта, составляющими те или иные подходы в рамках одной из трёх парадигм.

Что интересно, если попытаться расположить в матрице подходы и парадигмы в строках, а решаемые задачи в столбцах, то получится своеобразная «периодическая система технологий искусственного интеллекта», в ячейках которой будут перечислены различные методы конкретного подхода для решения конкретной задачи. Вот так может выглядеть такая матрица.

		Поиск	Обработка ЕЯ	Представ- ление знаний	Машинное обучение	Распозна- вание образов	Дата- майнинг	НЕ- факторы	Принятие решений	Робото- техника	Роевой интеллект
Нисходящая парадигма	Интуитивный подход	Тест Тьюринга Tt								Расширенный Тест Тьюринга Xt	
	Логический подход	Is Поиск информации	Mm Модель Маркова	Pr Продукционная модель	De Дедуктивное обучение	Pm Сопоставление с образцом	Rg Регрессионный анализ	Ds Теория Демпстера - Шефера	Gs Универсальный решатель задач	Rb На основе правил	Au Автоматы
	Символьный подход	Ss Поиск в пространстве состояний	Fg Формальные грамматики	Sn Семантические сети	Kb Базы знаний	Sf Семантическая свёртка	Dt Деревья решений	Fl Нечёткая логика	Es Экспертные системы	Fr Фреймы	Li Языки взаимо- действия
Гибридная парадигма	Агентный подход	Rw Случайное блуждание			Rl Обучение с подкрепле- нием				Cm Кибернети- ческая машина	Ro Роботы	Ra Рациональные агенты
Восходящая парадигма	Структурный подход		Sm Стат. методы обработки ЕЯ		An* Искусственные нейронные сети		Sd Стат. методы дата-майнинга				
	Эволюционный подход	Ga Генетические алгоритмы			Ne Нейро- эволюция		Ep Эволюционное программи- рование				Al Искусственная жизнь
	Квазибиологиче- ский подход	Dc ДНК- компьютер								Be Био- электроника	Nb Наноботы
* Искусственные нейронные сети		Pc Перцептроны	Wn Сеть Ворда	Bm Машина Больцмана	Db Глубокая сеть доверия	Km Карта Кохонена	Ae Авто- кодировщики	Nt Нейронная машина Тьюринга	Fn Свёрточные нейросети	Gn Генеративно- состязательные сети	DI Глубинное обучение

Периодическая система технологий искусственного интеллекта

Далее в этой книге во второй главе описано большинство представленных методов решения типовых задач искусственного интеллекта. Ну а пока рассмотрим каждый из перечисленных подходов к построению искусственных интеллектуальных систем подробнее.

Интуитивный подход к искусственному интеллекту был предложен Аланом Тьюрингом в своей ставшей уже знаменитой статье «Вычислительные машины и разум», опубликованной в 1950 г. В ней он предложил процедуру, которая, по его мнению, должна определить на интуитивном уровне, обладает некоторая система интеллектом или нет. Впоследствии эта процедура получила наименование «тест Тьюринга», подвергалась многочисленной критике, была расширена для систем с богатым набором сенсоров и, в общем, дожила до наших дней в качестве вполне себе операционной процедуры. Тем не менее сам подход не является конструктивным, поскольку Тьюринг нигде не говорил о том, *как* построить интеллектуальную систему.

Честно говоря, сам Тьюринг изложил свой тест немного путано (да ещё и в нескольких вариантах), так что сегодня исследователи ломают копыта на тему того, каким образом интеллектуальная система должна проходить данный тест. Но большинство исследователей сходится в одном: прохождение системой теста Тьюринга – необходимый фактор для того, чтобы признать за системой наличие интеллекта, но отнюдь не достаточный. Впрочем, и с этим утверждением многие будут спорить, так как сама природа человеческого интеллекта до сих пор не ясна, потому и бессмысленна разработка какого-либо теста для проверки на наличие интеллекта, «похожего на человеческий». Тем более что основная критика интуитивного подхода заключается в том, что моделирование естественного интеллекта не является единственной возможностью создания интеллектуальной системы. Искусственный интеллект может быть построен на совершенно иных принципах, как это ранее бывало со многими иными изобретениями (колесо для передвижения, реактивная тяга для полёта по воздуху, радиоволны для систем связи и т. д.).

Тем не менее, несмотря на то что тест Тьюринга и интуитивный подход в целом не могут считаться серьёзными инструментами в вопросе разработки искусственного интеллекта, сам по себе тест Тьюринга позволяет определить тот минимальный набор технологий и решаемых задач, которые должна обеспечивать система, чтобы считаться интеллектуальной. Кратко перечислим их.

1. *Обработка естественного языка*: интеллектуальная система должна уметь общаться с человеком на естественном языке, воспринимая все его неоднозначности, неопределённости и умолчания.

2. *Представление знаний*: в рамках искусственного интеллекта должны быть представлены как общие, так и специальные знания, при этом система должна постоянно обучаться и пополнять свою базу знаний, в том числе и в процессе диалога с человеком.

3. *Логический вывод*: используя знания и получаемые на вход запросы от человека, интеллектуальная система должна осуществлять правдоподобный логический вывод, который позволяет сформировать ответ на том же естественном языке. 4. *Машинное обучение*: система искусственного интеллекта должна быть адаптивной и приспосабливаться к меняющейся ситуации в общении, используя имеющиеся у неё знания в качестве шаблонов и применяя их к схожим ситуациям и, само собой разумеется, актуализируя по результатам свои знания об окружающей среде.

5. *Дополнительные сенсоры и исполнительные устройства* используются в так называемом «полном тесте Тьюринга», в котором система искусственного интеллекта должна действовать в естественной среде обитания человека, воспринимая её при помощи таких же датчиков, какие есть у человека (видеокамеры, аудиосенсоры, газоанализаторы и др.), и воздействуя на среду при помощи разного рода манипуляторов.

Перечисленное уже наводит на размышления о том, что тест Тьюринга направлен на выявление того, что искусственный интеллект должен успешно «мимикрировать» под человека, хотя для наличия интеллектуальных способностей это совершенно не требуется. Ни одна из перечисленных технологий сама по себе не необходима для того, чтобы считать искусственный объект интеллектуальным. Но в целом наличие этих пяти пунктов с прохождением полного теста Тьюринга позволяет говорить о том, что система *может* иметь интеллект, похожий на человеческий.

Логический подход основывается на формальной логике. Ведь ещё древнегреческий философ Аристотель сделал успешную попытку формального описания законов человеческого мышления. Сложно сказать, насколько эти законы универсальны, ведь кроме человеческого мышления у нас нет никаких иных примеров, но общее осмысление данного вопроса подсказывает, что формальная логика относится к чистому математическому знанию, т. е. находится в мире чистых идей, а потому *может быть* общезначимой. Другими словами, мышление, основанное на формальной логике Аристотеля, может быть универсальным, а потому реализация логических правил в искусственной системе может сделать её интеллектуальной. Однако тут имеется та же самая ловушка, как и в случае интуитивного подхода к искусственному интеллекту. Логическое мышление является необходимым условием, но никак не достаточным. Система, обладающая разумом, будет показывать поведение, подчиняющееся законам формальной логики. Но если какая-либо система действует по этим законам, это совсем не значит, что она интеллектуальна.

В рамках математики разработано большое количество формализмов, описывающих логику. Двоичная логика Аристотеля является базовой, над которой надстроены такие варианты, как многозначная логика Лукасевича, нечёткая логика Заде, бесконечнозначная логика антиномий, интуиционистская логика и некоторые другие. Каждый новый формализм был разработан для того, чтобы учесть какие-либо нюансы человеческого мышления и способа принятия решений человеком. Ведь в процессе исследований в направлении логического подхода

становилось понятным, что двоичная логика, хотя и является универсальным инструментом размышления и вывода знаний, плохо справляется с такими простейшими аспектами человеческого поведения, как принятие решений в условиях неопределённости, неполноты знания, неточности измерений и т. д.

Однако логический подход сталкивается с серьёзными трудностями, когда возникает необходимость описания неформальных знаний, которые плохо формализуются. Более того, из-за неполноты нашего понимания природы человеческого мышления логический подход всё так же страдает от невозможности полностью описать процесс мышления и принятия решений. Такие феномены, как озарение, интуитивный поиск решения или эмоциональные влияния на принятие решений, не могут быть описаны в рамках логического подхода, хотя они, вне всяких сомнений, являются одним из компонентов человеческого разума.

Тем не менее логический подход составляет основу упомянутой ранее нисходящей парадигмы искусственного интеллекта.

Символьный подход ставит во главу искусственного интеллекта способность человека манипулировать символами (в общем понимании этого термина) при осуществлении своей интеллектуальной деятельности. Тут нужно вспомнить термин «вторая сигнальная система», предложенный русским физиологом Иваном Петровичем Павловым для описания абстрактной системы обозначений, которые используются при мышлении человеком в отрыве от непосредственных ощущений, получаемых всеми сенсорными подсистемами нервной системы. Именно наличие второй сигнальной системы, по мнению некоторых исследователей, отличает разумное существо от животного.

Для упрощения рассмотрим кибернетическую цепочку возникновения реакции человека на стимул из внешней среды на примере зрения. Фотоны различной длины волны попадают на рецепторы сетчатки глаза, где запускают каскад биохимических реакций, результат которых заключается в генерации нервного импульса, идущего по главному нерву в нервные центры в головном мозге человека. Далее этот импульс диспетчеризуется в ядрах зрительного перекрёста и таламусе, после чего попадает в различные зоны коры головного мозга – первичную зрительную кору, вторичную зрительную кору и т. д., пока след от этого импульса не будет обработан в высших слоях неокортекса, где, как предполагается, и осуществляется манипуляция символами. Именно здесь конкретные каскады биохимических реакций в клетках сетчатки и всех промежуточных нейронах каким-то образом преобразуются в *абстрактные символы*, имеющие огромное количество ассоциативных связей. И когда человек видит, скажем, кошку, то в его памяти возникают различные образы, именно на символическом уровне связанные с понятием «*Feliscatus*» («кот домашний», биологическое наименование вида). Человек осознаёт именно эти образы и ассоциативные связи, но никак не осознаёт лежащих в их основе электрофизических и биохимических процессов. Так что разум человека, его интеллект оперирует именно символом «кот домашний», пробегая по отдельным ассоциативным связям, выбор которых зависит от текущего контекста.

Обработка символов построена на выполнении правил различного вида. В дальнейшем мы рассмотрим несколько формализмов, которые конкретизируют и реализуют символический подход к искусственному интеллекту. Однако все они основаны именно на манипулировании символами как синтаксическими конструкциями. Особенно это касается математических формул, поскольку наибольших успехов символический подход достиг именно в формальной математике, в таких областях, как автоматическое доказательство теорем, символическая математика, автоматические вывод и рассуждения и т. д. Однако и тут возникает такая же возможность для критики, как и в случае интуитивного подхода, – где возникает сознание, в какой момент интеллектуальная система начинает осознавать себя. И как бы хорошо она ни манипулировала символами, ответить на этот вопрос на сегодняшний день возможным не представляется. Мы

ещё вернёмся к этому аспекту в главе про философию сознания и искусственного интеллекта, а пока перейдём к следующему подходу.

Структурный подход, или *коннекционизм*, исходит из понимания того, что интеллект, разум и сознание являются функцией сложности сети переплетённых и взаимодействующих базовых элементов. Поскольку нервная система человека состоит из нейронов и ряда других специальных клеток, то основой структурного подхода в искусственном интеллекте является использование искусственных нейронных сетей.

Здесь уже упоминалось, что первый искусственный нейрон был разработан Уорреном Мак-Каллоком и Уолтером Питтсом в первой половине XX века. Также в то время Фрэнком Розенблаттом была разработана первая архитектура искусственных нейронных сетей – перцептрон. Фактически это самые первые примеры математических моделей и программных систем, разработанных в духе коннективизма. И здесь необходимо отметить, что в процессе развития перцептронов появились некоторые расширения первоначальной модели, предложенной Ф. Розенблаттом. Самый простой классификатор основан на количестве слоёв в перцептроне: однослойный, с одним скрытым слоем (классический) и многослойный. Все эти типы были в своё время описаны самим Ф. Розенблаттом.

Другая классификация включает: элементарный перцептрон, простой перцептрон, перцептрон с последовательными связями, перцептрон с перекрёстными связями, перцептрон с обратными связями, перцептрон с переменными связями. Первые три класса были описаны Ф. Розенблаттом, а следующие три развиты в дальнейшем при детальной проработке модели искусственных нейронных сетей.

Итак, структурный подход основан на идее о том, что наиболее важной для эмерджентного проявления разума и других подобных функций вплоть до сознания является сетевая структура. Это значит, что моделирование должно затрагивать не только базовые элементы типа нейронов, но и их взаимосвязанные сети. При этом полагается, что сеть элементов как сложная система обладает нелинейными свойствами относительно базовых элементов, а потому при увеличении количества взаимосвязанных элементов сложность модели изменяется нелинейно и зачастую даже непредсказуемым образом, так что при переходе некоторого порога сложности как раз и начинают проявляться те самые эмерджентные эффекты, которых мы так ждём. До сегодняшнего дня главным и фактически единственным представителем структурного подхода являются искусственные нейронные сети.

Эволюционный подход реализует так называемую «искусственную эволюцию» и фактически решает оптимизационную задачу поиска значения целевой функции в заданном пространстве решений с установленными ограничениями. Этот подход содержит большое количество эвристических методов, некоторые из которых используются даже для гарантированного нахождения оптимума для мультимодальных недифференцируемых функций в пространствах высоких размерностей. Важным методом эволюционного подхода являются генетические алгоритмы, которые мы рассмотрим чуть позже. Но здесь также есть несколько других важных методов, в числе которых находятся эволюционное программирование и даже нейроэволюция, когда искусственному отбору подвергаются нейронные сети.

Итак, в рамках эволюционного подхода обычно выделяют следующие технологии и методы: эволюционное программирование, генетическое программирование, эволюционные стратегии, генетические алгоритмы, дифференциальная эволюция и нейроэволюция. Другими словами, а что, если вычислительные процессы могли бы эволюционировать так же, как это делают биологические виды в своей экологической среде? Возможно, получилось бы «выращивать» программы для оптимального решения поставленной задачи? Эволюционное программирование как раз и основано на этой идее.

Если же в качестве объектов отбора выступают сами программы, то получается уже генетическое программирование. Ведь, действительно, программы пишутся на определён-

ном языке программирования и в конечном итоге представляют собой строки символов. Эти строки можно подвергнуть генетическим преобразованиям и отбору. Эта очень мощная идея получила своё развитие в том, что программы начали писать и оптимизировать другие программы, и уже исследователи, запустившие процесс, не могут разобрать и интерпретировать полученные исходные коды, которые работают правильно и часто очень эффективно.

Генетические алгоритмы – это наиболее яркий представитель эволюционного подхода. Сами по себе они опять являются одним из эвристических методов оптимизации для поиска оптимального решения (или, как минимум, субоптимального). Они работают с данными, которые могут быть представлены в виде «хромосом» – последовательностей генов, т. е. списков каких-либо значений, к которым можно применить генетические операции. Здесь главное – чтобы на генах были определены эти самые генетические операции, которые возвращали бы приемлемый результат, имеющий смысл.

Мы ещё детально ознакомимся с эволюционными алгоритмами и вообще эволюционным подходом во второй главе.

Идём дальше и кратко коснёмся **квазибиологического подхода** и его методов. Фактически это отдельное направление исследований в искусственном интеллекте, так как оно основано не на цифровом моделировании *in silico* разных аспектов интеллекта, а на применении биомолекулярных механизмов для того же самого. Этот подход ещё называется «биокомпьютингом», и это очень перспективное направление. В рамках квазибиологического подхода разработано большое количество методов – начиная от биомолекулярной электроники, молекулярных вычислений и заканчивая нейрокомпьютингом. В последнем направлении важной вехой является разработка нейроморфных чипов. И иногда мне кажется, что прорыв в области искусственного интеллекта будет лежать на пересечении трёх областей – техники, информатики и химии. Но посмотрим...

В основе этого подхода лежит понимание, что феномены человеческого поведения, наша способность к обучению и адаптации есть следствие именно биологической структуры и особенностей её функционирования. Хотя, скорее всего, это очень слабая гипотеза. Вычисления в рамках квазибиологического подхода организуются при помощи живых тканей, клеток, вирусов и различных биомолекул. Часто используются молекулы дезоксирибонуклеиновой кислоты, на основе которой создают ДНК-компьютер. Кроме ДНК, в качестве биопроцессора могут использоваться также белковые молекулы и биологические мембраны.

Обычно для решения определённой задачи создаётся так называемая «индивидуальная машина», которая, в отличие от универсальной машины Тьюринга, направлена на решение конкретной задачи, причём обычно делает это более эффективным способом, поскольку индивидуальная машина специально сконструирована для решения именно этой задачи. Машина Тьюринга, лежащая в основе стандартной вычислительной модели, выполняет свои команды последовательно, а в рамках квазибиологической парадигмы часто рассматривается массовый параллелизм. Ну вот если, к примеру, рассмотреть ДНК-компьютер, то в нём все молекулы ДНК одновременно участвуют во взаимодействиях, параллельно проводя вычисления.

Два самых главных направления в рамках квазибиологического подхода – это молекулярные вычисления и биомолекулярная электроника. Можно ещё упомянуть нейрокомпьютинг и создание нейроморфных чипов, но они чаще всего рассматриваются как часть структурного подхода и искусственных нейронных сетей.

Молекулярные вычисления – это отдельная вычислительная модель, в которой решение задачи осуществляется при помощи проведения сложных биохимических или нанотехнологических реакций. Молекулярные компьютеры – это молекулы, запрограммированные на нужные свойства и поведение, которые, участвуя в химических реакциях, как бы «вырачивают» результат. Что интересно, идею биокомпьютинга подсказал выдающийся математик Джон фон Нейман в своей книге «Теория самовоспроизводящихся автоматов», которую,

кстати, очень рекомендую для внимательного чтения. В этой книге описан проект клеточных автоматов, которые могут самовоспроизводиться, как живая клетка.

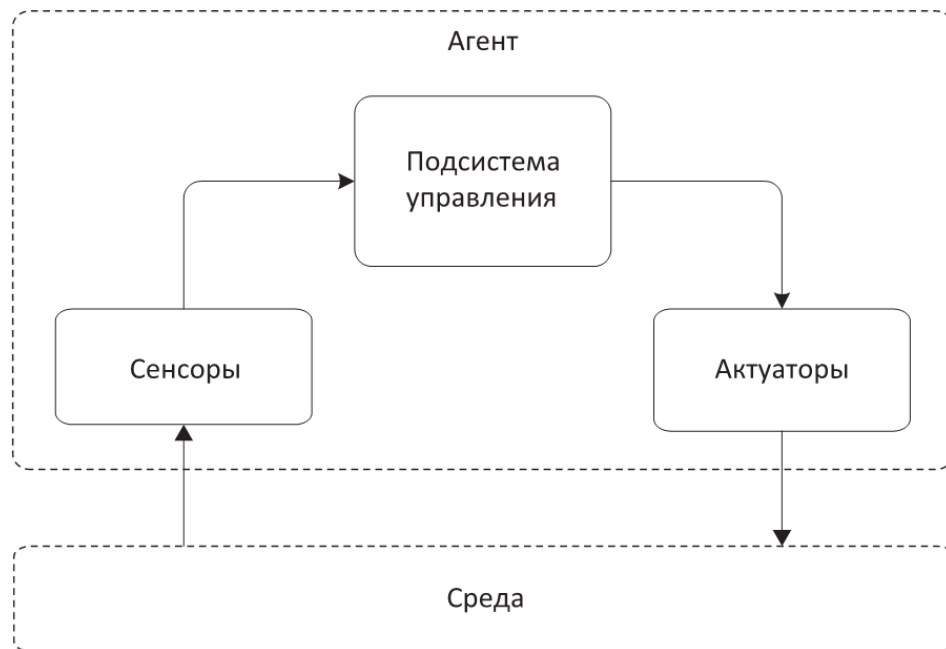
Почти в каждой живой клетке нашего организма есть длинная молекула ДНК, кодирующая генетическую информацию. При помощи различных ферментов цепочки ДНК могут быть разрезаны, склеены, в них могут добавляться буквы генетического кода или удаляться из них. Всё это – базовые операции работы с информацией, которые могут быть использованы для производства вычислений. Более того, цепочки ДНК могут воспроизводиться и клонироваться. Это позволяет запустить массовый параллелизм поиска решения. В небольшой пробирке после проведения должным образом сконструированной биохимической реакции будет получен результат, который считывается специальной аппаратурой.

Интерес вызывает то, что для некоторых задач молекулярные компьютеры очень быстро и точно находят приемлемые решения, в то время как традиционные компьютеры затрудняются это сделать. Например, решение задачи коммивояжёра, т. е. поиска кратчайшего пути обхода графа, при помощи реакций с ДНК осуществляется практически мгновенно, в то время как для обычного компьютера требуется огромное количество времени. Правда, тут есть одна тонкость, которая мешает работе обычному компьютеру, – это комбинаторный взрыв. И если в традиционной архитектуре он ведёт к увеличению времени решения, то для ДНК-компьютера требуется подготовка огромного количества вариантов нуклеотидных нитей. Соответственно, объём пробирки растёт так же, как и количество вариантов в комбинаторном взрыве.

В общем, часто биокомпьютинг можно охарактеризовать как новую парадигму вычислений, которая, в отличие от традиционной вычислительной модели, работает быстро, но при решении сложных задач с комбинаторным взрывом растёт не время вычислений, а необходимый для них объём биокомпьютера.

Вместе с тем в последнее время всё активнее разрабатывается **агентный подход** к построению искусственного интеллекта. В рамках этого подхода изменена точка зрения на цель построения интеллектуальной системы и считается, что построить нужно систему не с разумным поведением, а с рациональным. С одной стороны, это серьёзно облегчает задачу, поскольку, в отличие от понятий «разум» или «интеллект», понятия «рациональность» и «рациональное поведение» можно строго формализовать (например, рациональное поведение – это выбор и достижение оптимальной цели с минимизацией затраченных на это ресурсов). С другой стороны, для демонстрации рационального поведения агент должен обладать достаточной «разумностью», чтобы определить цель, составить стратегию её достижения и выполнить её.

Каждый агент – это полноценная кибернетическая машина, которая имеет систему управления, непрерывно получающую информацию с сенсорных систем агента и воздействующую на окружающую среду при помощи исполнительных устройств (или *актуаторов*). При этом подход не определяет сущность сенсорных систем и актуаторов – их природа может быть произвольной. Поэтому агентный подход одинаково применим как к чисто программным сущностям, работающим в некоторой искусственной среде, так и к программно-аппаратным комплексам, равно как и вообще к биологическим системам.



Общая схема агента и его взаимодействия со средой

Агентный подход интересен тем, что в его рамках можно использовать эволюционные алгоритмы, которые подбирают интеллектуальных агентов, исходя из степени их приспособленности к достижению цели. Во время взаимодействия агентов осуществляется отбор наиболее успешных, которые затем используются для генерации нового поколения агентов, среди которых опять применяются те же самые процедуры оценки и отбора. В итоге наиболее успешное поколение решает задачи и достигает целей наиболее эффективным образом. Это идеальный вариант, который сегодня сложно достижим, но стремиться к нему интересно. Также агентный подход лежит в основе так называемых многоагентных систем, в рамках которых осуществляется общее целеполагание, после чего каждому индивидуальному агенту даётся свобода действий в определённых рамках, где он имеет возможности и альтернативы по разработке и реализации различных стратегий достижения своей частной цели. В процессе этого агенты взаимодействуют друг с другом и со средой, обмениваясь информацией и выполняя запросы других агентов. Кроме того, вполне может быть использована идея так называемого «роевого интеллекта», когда каждая отдельная «особь» (то есть агент) интеллектом не обладает, но в целом «рой» (множество агентов, многоагентная система) обладает определёнными интеллектуальными способностями. Надо отметить, что всё перечисленное является одним из наиболее перспективных направлений исследований по искусственному интеллекту.

* * *

Итак, философия сознания, начавшаяся развиваться вместе с более «взрослым» пониманием задач и проблем искусственного интеллекта, выделяет два крупных подхода к построению интеллектуальных систем и искусственных разумных существ: чистый (нисходящий) и грязный (восходящий). Чистый подход объединяет такие технологии, как экспертные системы, универсальные машины вывода, семиотические базы знаний. Самым выдающимся и далеко продвинувшимся направлением в рамках этого подхода является направление символьных вычислений, которое основано на логике манипулирования символами. Грязный подход объединяет такие технологии, как искусственные нейронные сети, эволюционные вычисления и биокомпьютинг. Грязный подход моделирует биологические основы разума у человека, в то

время как чистый подход имитирует высокоуровневые психические и когнитивные процессы: мышление, рассуждение, речь, эмоции, творчество и т. д.

Однако, как уже было описано выше, и чистый, и грязный подходы имеют как достоинства, так и свои недостатки. В частности, нейронные сети и все смежные технологии обладают очень существенными ограничениями.

1. Наука до сих пор не обладает полным пониманием того, как обучается нейронная сеть. Теоретический математический аппарат, конечно, имеется, однако при переходе в практическое русло вычислительная сложность интерпретации того, как и, главное, почему нейросеть настроила свои весовые коэффициенты так, а не иначе, нелинейно возрастает до космических величин. И в итоге получается чёрный ящик в качестве модели чёрного ящика. Это совсем не то, что нужно.

2. Часто в нейронную сеть для обучения попадают некорректные данные, которые выглядят как валидные. Например, в медицинской проблемной области нередко случается так, что диагнозы ставятся неверно, лекарственные средства назначаются неадекватно, а потому нет никакой гарантии того, что при обучении нейросети ей не «скармливаются» такие некорректные данные. Принцип GIGO никто не отменял. Кто даст гарантию того, что обученная на некорректных данных нейросеть даст правильное заключение, а инструмента интерпретации того, как она обучилась, нет.

3. В процессе развития человеческих знаний всегда появляется новая информация, которая часто не дополняет, а заменяет старую. Человек-эксперт постоянно актуализирует свои знания, занимаясь самообразованием и повышением квалификации. Как поступать с нейросетью? Ведь невозможно проинтерпретировать то, как она обучилась, а потому нет никакого способа точно заменить устаревшую информацию. Сеть приходится переобучать с нуля.

4. Более того, «углеродная» нейросеть фундаментально отличается от «кремниевой» – сегодня ещё не разработано окончательной рабочей модели того, как обучается человек. С учётом непрерывающегося нейрогенеза в отдельных областях головного мозга есть понимание, что в нейросетях в голове человека могут появляться не только новые отдельные нейроны, но и даже целые слои. Впрочем, и нейроны, и слои также могут исключаться из процесса вычислений. И если появление и исключение синаптических контактов можно эмулировать весовыми коэффициентами в искусственных нейронных сетях, то про искусственные модели с появлением новых слоёв пока слышно очень редко и тихо. 5. К тому же центральная нервная система человека представляет собой не обычную многослойную нейросеть, а десятки отдельных сетей, связанных друг с другом. Иногда связанных совсем нелинейным образом – с кучей реципрокных и обратных связей, которые могут выходить и входить в промежуточные слои любой сети. Сигнал от сенсора до исполнительного устройства в человеке может пройти через десяток коммутирующих ядер, запустить параллельные процессы и в итоге так и не дойти до неокортекса, оставшись в подсознательной области.

6. Наконец, что касается неокортекса, где, как предполагается, и скрывается тайна сознания. Обработка информации в нейросети неокортекса может завершиться в любом из шести его слоёв, но только, как считают некоторые исследователи, если сигнал дойдёт до пятого или шестого слоя, он будет доступен для осознания при помощи символической логики. Вся подсознательная деятельность (можно сказать, что она лежит в основе интуиции) готовит базу для принятия решения на основе механизма манипулирования символами. Именно этот аспект надо рассмотреть более подробно.

Итак, вся накопленная человечеством информация представлена в символическом виде. И именно логическая (синтаксическая) манипуляция символами позволяет делать какие-либо умозаключения. Это совершенно иной подход, фундаментально отличающийся от нейронных сетей, какими бы глубокими они не были. Вполне вероятно, что где-то там в очень глубокой нейросети может зародиться логика манипулирования символами, но пока это неизвестно в

силу уже неоднократно упомянутого отсутствия формализма интерпретации работы нейросетей. А с другой стороны, математика предоставляет достаточное число формализмов и инструментов для логической обработки символов.

Символьный подход в искусственном интеллекте сегодня незаслуженно отставлен в сторону. Скорее всего, это произошло из-за раннего разочарования в нём по причине того, что естественный язык как наиболее развитая символьная система, используемая человеком, является очень сложной и плохо формализуемой. Неопределённость высказываний, нечёткость формулировок, разного рода неточность измерений параметров – всё это вносит вклад в повышение сложности принятия решений на основе работы с символьной информацией. И если человек как-то справляется с этим в условиях неопределённости, неполноты и даже противоречивости входной информации, то обучить этому искусственную интеллектуальную систему достаточно сложно. Впрочем, это возможно.

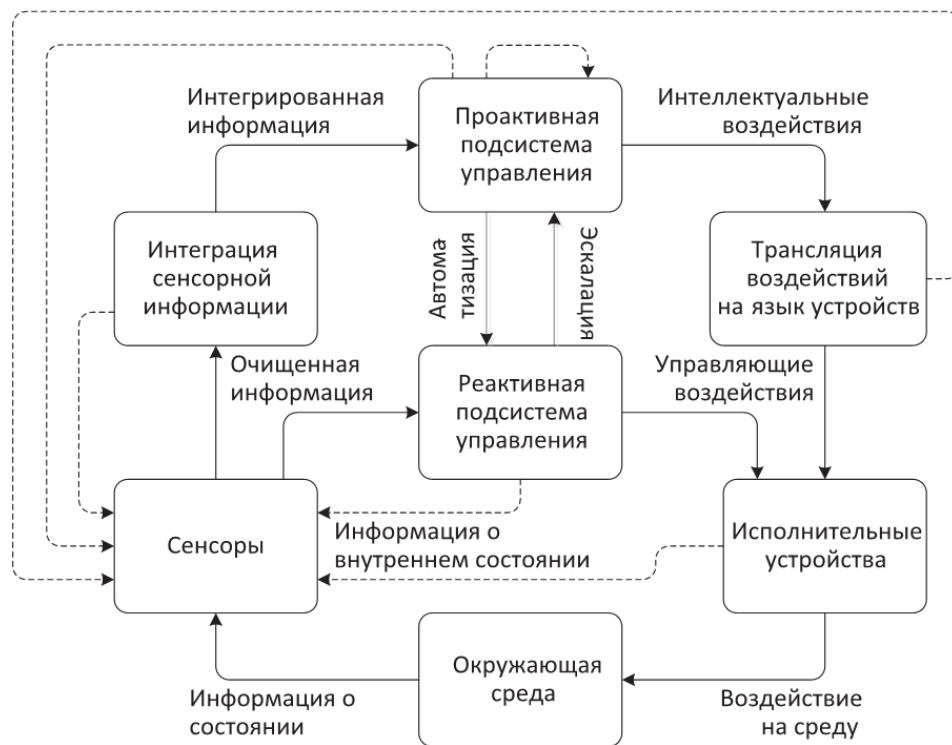
Основная проблема символьного подхода заключается в быстро достигаемом «комбинаторном взрыве» при попытках разобраться в семантике обрабатываемых формул. Большое количество синонимов, нечёткие лингвистические переменные с не до конца определёнными шкалами – всё это отпугивало экспертов и повергало в уныние инженеров. Для того чтобы создать базу знаний для самой простой проблемной области, приходилось либо записывать тысячи правил на все случаи жизни, либо постоянно повышать уровень абстракции представления до того уровня, когда это представление становилось неадекватным решаемым задачам. В качестве примера можно привести базу знаний по выдаче предположений о возможных заболеваниях по общему анализу крови – база содержит более 1000 продукций, но при этом всё ещё не может охватить все возможные случаи, встречаемые в практике.

Однако сегодня описанные проблемы вполне поддаются решению. Из-за того же роста вычислительных мощностей, который помог развитию структурного подхода, вполне можно возобновить работы в рамках символьного направления искусственного интеллекта. Во многих проблемных областях знания в подавляющем большинстве случаев либо формализованы, либо хорошо поддаются формализации. Главной задачей становится создание нормальной машины вывода, которая справится с «комбинаторным взрывом».

Уже описанный символьный подход в искусственном интеллекте позволяет решить главную проблему, возникающую при использовании искусственных нейронных сетей и эволюционных алгоритмов, – интерпретацию результатов. Структурированные и формализованные знания и вывод на них позволяют осуществить пошаговое объяснение того, как был получен тот или иной результат. Все остальные особенности структурного подхода вполне доступны и в рамках символьного.

И тут хотелось бы отметить, что рождение искусственного интеллекта, скорее всего, произойдёт где-то на стыке двух подходов, т. е. при использовании **гибридной парадигмы**. Вряд ли в ближайшее время удастся реализовать *in silico* аналог человеческой нервной системы, поскольку её сложность невообразима. Вероятно, проще вырастить биологическую нейросеть такого объёма, чем найти вычислительные мощности для её симуляции. Однако нейросетевой подход даст базовые инструменты для первичной обработки и коммутации информации, поступающей с разнообразных сенсоров на вход системе управления и принятия решений в составе искусственного интеллекта. А в ней уже будет использоваться символьный подход, реализованный в виде универсальной машины вывода. И такая конвергенция двух технологий в конце концов позволит реализовать давнюю мечту человека – рождение искусственного разума.

Следующая диаграмма графически показывает общую схему взаимодействия компонентов в гибридной интеллектуальной системе, которая основана на описанных ранее принципах.



Итак, универсальная машина вывода, или подсистема управления, теперь разбита на две части.

1. *Реактивная подсистема управления* фактически реализует традиционную схему управления, когда сигналы с сенсоров обрабатываются системой управления и по ним осуществляется формирование управленческих воздействий на среду (объект управления) через исполнительные устройства.

2. *Проактивная подсистема управления* добавляет дополнительный промежуточный уровень, который позволяет осуществлять обучение системы, построение прогноза на основе моделирования среды и своего поведения в ней, построение плана действий и сравнение факта с прогнозом и планом для осуществления обучения (адаптации) системы к изменяющимся условиям внешней среды.

Эти подсистемы связаны друг с другом при помощи передачи *фокуса управления*. Когда проактивная подсистема создаёт новый паттерн поведения системы в изменившихся условиях и среда «устоялась», контур управленческого воздействия спускается в реактивную систему, происходит автоматизация реакции, так как фактически для неизменяющихся условий деятельности обучение и проактивное поведение не требуется, а потому реактивная реакция будет работать быстрее. Если же в процессе «рефлекторного» действия реактивной системы обнаруживается изменение в среде или объекте управления, то реактивная система эскалирует фокус внимания на проактивную для обработки изменившихся условий и выработки новых правил и паттернов поведения.

Цикл управления в рамках такой интеллектуальной системы управления теперь заключается в последовательном выполнении следующих шагов.

1. Сбор входной информации со всех сенсоров, которые осуществляют мониторинг различных параметров объекта управления и среды, в которой функционирует система. Каждый тип сенсора в этом случае является отдельной индивидуальной *модальностью восприятия* гибридной искусственной интеллектуальной системы.

2. Сенсоры очищают входную информацию от шумов и осуществляют первый выбор пути дальнейшей обработки. Если входная информация с сенсоров соответствует каким-либо авто-

матическим *паттернам поведения* системы, то фокус управления передаётся в реактивную подсистему, которая выбирает конкретный паттерн и исполняет его. Однако если в процессе реагировать по реактивному сценарию, система обнаруживает, что что-то пошло не так, осуществляется эскалация на проактивную подсистему управления, как было описано ранее.

3. Если входная информация не имеет автоматической реакции для своей обработки, то осуществляется интеграция всех модальностей восприятия системы в единый блок описания объекта управления и среды. На выходе этого модуля появляется целостная картина восприятия, которая передаётся в проактивную подсистему управления.

4. Проактивная подсистема управления принимает решение на основе имеющихся у неё динамических моделей себя самой, объекта управления и среды. Здесь как раз и используются методы машинного обучения и нисходящей парадигмы искусственного интеллекта. На выходе проактивной подсистемы управления появляется управленческое воздействие, которое записывается в реактивную подсистему в качестве нового правила, а также направляется на исполнение.

5. Для исполнения управленческое воздействие переводится на язык конкретных исполнительных устройств, которые взаимодействуют с объектом управления и средой. Исполнительные устройства выполняют команду. Цикл работы завершается.

Таким образом видно, что гибридная интеллектуальная система отличается тем, что её аффлекторы (сенсоры, датчики) и эффлекторы (исполнительные устройства) связаны с подсистемой управления и принятия решений через нейронные сети. Тем самым реализуется «грязный» подход. Аффлекторная нейронная сеть принимает очищенные сенсорами сигналы внешней среды и преобразует их в символы, которые подаются на вход универсальной машины вывода. Последняя осуществляет вывод на основе символьных знаний из своей базы знаний и выводит результат, который тоже представляется в виде символов. Тем самым реализуется «чистый» подход. Символьный результат подаётся на вход моторной нейронной сети, которая преобразует высокоуровневые символы в конкретные сигналы управления исполнительными устройствами.

Кроме всего прочего, внутри гибридной интеллектуальной системы должны быть реализованы контрольные связи от всех её элементов к сенсорам. Тем самым реализуются адаптационные механизмы, основанные на гомеостазе внутреннего состояния системы. Сенсоры фиксируют изменение внутреннего состояния каждой подсистемы, их элементов и комплексов, а в случае выхода контролируемых значений за пределы установленных гомеостатических интервалов подсистемой управления принимается решение, целью которого будет возврат изменённых показателей в установочный интервал.

Наконец, связь от проактивной подсистемы управления в саму себя олицетворяет так называемый «внутренний конфликт», когда искусственная интеллектуальная система может рассматривать и моделировать различные варианты развития ситуации. Цикл оценки и выбора приемлемой альтернативы осуществляется до тех пор, пока этот внутренний конфликт не будет улажен.

Интерес вызывает то, что система именно с такой архитектурой при переходе через определённый порог сложности может считаться разумной. Действительно, в соответствии с определением, данным нейрофизиологом Е. Р. Джоном (цитируется по книге Хофштадтера и Деннета «Глаз разума»), сознание – это «процесс, в ходе которого информация о множественных индивидуальных модальностях восприятия и ощущения сводится в единое многоплановое представление о состоянии системы и её окружения и интегрируется с информацией о воспоминаниях и потребностях организма, порождая эмоциональные реакции и программы поведения, способствующие приспособлению организма к его окружению». Представленная на предыдущей диаграмме кибернетическая схема на довольно высоком уровне абстракции описывает, как функционирует нервная система человека и высших животных. Ведь «разум-

ность» в этом смысле определяется как адекватное реагирование не только на стимулы внешней среды, но и на внутренние состояния, что также включает в себя постоянный мониторинг состояния собственной подсистемы управления, что называется «саморефлексией», которая и приводит к осознанию.

* * *

Чтобы более точно понять изложенное, имеет смысл рассмотреть несколько примеров. Самым тривиальным примером интеллектуальной системы является человек. На определённом уровне абстракции человек как система выглядит именно так, как представлено на диаграммах выше. Роль сенсоров выполняют не только такие общеизвестные рецепторные системы, как зрение, слух, вкус, обоняние, осязание и равновесие, но и различного рода ноцицепторы, терморецепторы, проприорецепторы и хеморецепторы. Именно последние из перечисленных осуществляют мониторинг гомеостаза внутренних параметров организма, начиная от кислотности крови и заканчивая наличием и концентрацией в физиологических жидкостях организма определённых веществ. Каждый тип рецепторов непрерывно воспринимает сигналы из внешней по отношению к нему среды, при этом сразу же осуществляется первичная фильтрация входных значений по полосе принимаемого диапазона. Например, фоторецепторы на сетчатке глаза воспринимают фотоны только определённых длин волн.

Далее отфильтрованные сигналы «оцифровываются» и передаются в первичные сенсорные нейронные сети, где осуществляется перекоммутация и задействование следующих областей нервной системы. Потом сигнал идёт во всевозможные центры принятия решений, и часто он даже не доходит до неокортекса, «поглощаясь» в подсознательных областях (но там также накапливается информация, которая в какой-то момент может стать основой и источником интуитивного инсайта). Если же сигнал доходит до неокортекса, в нём он начинает активно преобразовываться и задействовать множество ассоциативных связей, возбуждая побочные цепи. В итоге до верхних слоёв неокортекса префронтальных долей, где, как предполагается, и находится сознание, доходит символьная информация.

Например, если последовательность множества сигналов, падающих на сетчатку глаза, формирует на ней изображение движущейся рыжей кошки, то на верхних слоях неокортекса активированы те нейроны, которые отвечают за такие символы, как «кошка», «млекопитающее», «рыжее животное» и множество других.

Путём логических умозаключений (на самом деле последовательного возбуждения каскадов нейронов в ассоциативно-теменном и префронтальном участках неокортекса) формируется ответная реакция. Она также кодируется в виде символов. Например, человек, видящий кошку, может выбрать из всего множества ассоциаций одно слово «кошка», сосредоточившись на нём и подав сигнал нижележащим подсистемам активировать поведенческую реакцию типа «воскликнуть наименование объекта». Моторная нейронная сеть воспринимает этот символьный сигнал и начинает распространять его всё ниже и ниже. На самом нижнем уровне символ «кошка» преобразуется в динамическую последовательность сокращений мышц в разных отделах тела, чья совместная деятельность заключается в таком синхронизированном выполнении ряда элементарных сокращений, которые приведут к возникновению исходящего из лёгких потока воздуха с одновременным формированием в голосовых связках, нёбе, языке и зубах быстрого ряда сжатий и растяжений, приводящих к возникновению вибраций определённой амплитуды и частоты. Которые, в свою очередь, воспринимаются окружающими людьми и самим человеком, произнёсшим слово, как слово «кошка». Сам человек воспринимает ушами сказанное самим собой слово, которой через такой же каскад реакций и перекоммутаций нейронных сигналов доходит до контролирующего элемента неокортекса, что сравнивает заказанное и исполненное слово – они должны совпасть. Эта петля обратной связи контроля поведе-

ния широко используется в большинстве поведенческих актов человека. Всё это происходит практически мгновенно и без задействования сознания и внимания. Но похоже, что и сознание основано именно на этом механизме.

Другим примером является интеллектуальная система управления перекрёстком на улично-дорожной сети города. Здесь окружающей средой, в которой действует интеллектуальный агент, является дорожное движение. Сенсорами являются детекторы транспорта и другие периферийные устройства, которые снимают показатели транспортных и пассажиропотоков, значения выделенных погодных показателей и показателей локальной экологии, а также коммуникационное оборудование, принимающее сигналы операторов из центра управления. Кроме того, система постоянно опрашивает все подключённые к ней устройства на предмет их работоспособности, и это функция контроля собственного гомеостаза. Исполнительными устройствами являются светофоры, табло отображения информации и иные средства прямого и косвенного управления дорожным движением. В качестве подсистемы управления выступает контроллер светофорного объекта, выполняющий сценарии управления при возникновении тех или иных шаблонных или особых ситуаций на дороге, характеризующихся выделенными множествами значений входных параметров, включаются те или иные режимы управления дорожным движением.

Сенсорная и моторная нейронные сети здесь редуцированы до декодеров сигналов с оборудования и преобразователей сигналов верхнего уровня из подсистемы управления в замыкания конкретных электрических цепей, зажигающих лампочки или светодиоды в светофорах или на табло. Отличительной особенностью такого интеллектуального агента является то, что он может действовать автономно, а также он может взаимодействовать с другими агентами, входящими в состав интеллектуальной транспортной системы. При этом подобное взаимодействие направлено на кооперативное достижение единой цели оптимизации дорожного движения в целом путём «торга» во имя достижения локальных целей каждого конкретного интеллектуального агента.

* * *

Размышления над тем, может ли искусственная система мыслить так, как человек (что бы это ни значило – мы до сих пор не уверены относительно природы сознания), приводят к двум различным пониманиям искусственного интеллекта как практического направления:

- слабый искусственный интеллект;
- сильный искусственный интеллект.

Споры о том, что такое искусственный интеллект, начались практически сразу же после введения в научную практику определения нового понятия. Но до сих пор ответа на вопрос о том, что же такое «искусственный интеллект» и какие критерии можно использовать для того, чтобы определить, является ли искусственная система интеллектуальной и, шире, разумной, нет. С точки зрения философии сознания ответить на этот вопрос не так-то просто. Как уже было рассказано выше, первым разработать хотя бы какие-то операционные критерии попытался Алан Тьюринг в рамках интуитивного подхода. Попытка не очень удачная, тем не менее она вполне применима, хотя в дальнейшем были определены совершенно иные принципы построения интеллектуальных систем. И системы, построенные на этих принципах, вообще не могут быть подвергнуты тесту Тьюринга, хотя по иным критериям должны считаться интеллектуальными.

Масла в огонь подлил один из интереснейших философов современности Джон Сёрль, который предложил разделить понимание слабого и сильного искусственного интеллекта. Под сильным он понимал такой искусственный интеллект, который, если говорить самыми простыми словами, воспринимает самого себя в качестве отдельной личности. Это очень узкое

понимание, поскольку привлекает антропоморфное сравнение. Однако искусственное разумное существо не обязательно должно мыслить как человек. Поэтому говорить о том, что оно будет оперировать человеческими категориями, в том числе и при осознании себя, очень опасно с методологической точки зрения.

Но в итоге исследователи и разработчики интеллектуальных систем пришли к соглашению о том, что считать искусственным интеллектом. Под **слабым искусственным интеллектом** начали понимать интеллектуальные системы, которые решают слабо алгоритмизируемые или вовсе неалгоритмизируемые задачи методами, схожими с теми, которые использует для решения таких задач человек. Другими словами, слабый искусственный интеллект представляет собой реализацию отдельных когнитивных функций человеческого разума для достижения конкретных целей. При этом о наличии у таких систем самосознания, не говоря уже об иных высших функциях психологической деятельности человека, даже не говорится – их нет, и их существование даже не предполагается.

С другой стороны, **сильный искусственный интеллект** подразумевает появление у искусственной интеллектуальной системы самосознания с дальнейшим выходом в автономный режим жизнедеятельности и получением возможности перепрограммировать саму себя с целью усовершенствования. Именно с этим типом связывается точка достижения технологической сингулярности – сильный искусственный интеллект будет экспоненциально развиваться, поглощая всю накопленную человечеством информацию и используя её для собственного роста. Более того, он сможет генерировать новые идеи и концепции, как это делает человек, что в конце концов приведёт к тому, что сильный искусственный интеллект превзойдёт в интеллектуальных способностях всё человечество вместе взятое. Для обозначения таких искусственных разумных существ далее будет использоваться слово «ИскИн».

Наконец, в последнее время возникла новая парадигма в науке о сознании, которая пока ещё остаётся больше фантастической, нежели реальной. Речь идёт о так называемой **гибридизации сознания**, когда человеческий естественный интеллект сливается с искусственным интеллектом и при этом усиливается его методами и технологиями. Фактически цель этого направления исследований состоит в отыскании путей и возможностей для осуществления цифрового бессмертия человеческой личности, т. е. записи её брэнного биологического тела на цифровые носители. Это очень занятное направление исследований, которое находится в самом начале своего пути. Но тут есть многое, о чём можно подумать.

* * *

Необходимо отметить, что в конечном итоге работы над искусственным интеллектом в понимании искусственного разумного существа могут привести к его появлению. Сильный ИскИн может появиться, так как этому нет фундаментальных преград. Действительно, мы сами, люди, – отчётливый пример того, что разум, интеллект и сознание существуют в этой реальности, а это значит, что все понятия, описывающие конкретные явления, не являются умозрительными. Интеллект так или иначе существует, так что почему же надо отказывать себе в возможности его моделирования или даже создания? Абсолютно не стоит.

Однако размышления наводят на мысли о том, что, скорее всего, ИскИн может появиться в виде гибридной ИИИ-системы на стыке, как минимум, трёх дисциплин: информатики, химии с биологией и техники.



Действительно, описанная ранее квазибиологическая парадигма искусственного интеллекта сегодня является одним из мощно развивающихся направлений исследований. В чашках Петри уже научились выращивать искусственные нейронные сети на живых нейронах, которые обучаются управлять простыми механизмами. Бионические протезы и импланты, в том числе нейроимпланты, становятся реальностью на наших глазах. Достижения робототехники поражают сегодня даже выдавших виды инженеров и учёных.

Другими словами, информатика и весь комплекс наук в её основе даст понимание основ вычислительных процессов и того, как они моделируют психологические функции. Техника и все смежные дисциплины дадут возможность создавать физические тела для ИИ-систем. А химия и биология подведут к пониманию того, как в ИИ-системы встроить механизмы метаболизма и эволюции, что в конечном итоге даст ключ к пониманию жизни как таковой. На пересечении этих трёх достижений и родится ИскИн.

Однако всё может произойти и совсем не так...

На этой жизнеутверждающей ноте мы заканчиваем первую главу. Далее мы рассмотрим три столпа искусственного интеллекта и современное состояние всех технологий, которые сегодня объединяются термином «искусственный интеллект». Речь, конечно, о слабом его варианте. Так что начнём...

Глава 2

Три столпа искусственного интеллекта

Теперь давайте погрузимся в самые глубины методов и технологий искусственного интеллекта. Само собой разумеется, что далее на протяжении этой и следующей главы мы будем рассматривать только искусственный интеллект в его слабом варианте. Но именно это позволит нам немного раскрыть своё сознание и подготовиться к четвёртой главе, в которой мы рассмотрим разные интересные аспекты исследования разума и сознания, в том числе и при помощи методов искусственного интеллекта. Ведь надо помнить, что одной из первичных задач рассматриваемого нами междисциплинарного направления исследований является моделирование как отдельных когнитивных функций, так и всего разума, для того чтобы изучать его на модели. Это и есть главная задача – получить модель, чтобы познать природу человеческого разума.

Но пока давайте рассмотрим операционное определение Искусственного Интеллекта, которое позволит точно определять, когда об этом предмете говорят что-то стоящее, а когда пытаются развесить на уши лапшу в жёлтой прессе. Итак, искусственный интеллект – это междисциплинарная область исследований и набор технологий, позволяющий создавать технические системы, решающие задачи, ранее доступные только человеку. Ну вот как-то так.

И давайте теперь определим термин «искусственная интеллектуальная система», или проще – «ИИ-система». ИИ-система – это техническая система, которая при помощи методов и технологий искусственного интеллекта решает задачу или набор задач, ранее доступных только человеку. При этом ИИ-системы обладают двумя важными качествами, которые будут дифференцировать их в ряду других подобных систем. Это автономность и адаптивность, так называемые «2А».

- Автономность – это свойство системы, позволяющее ей действовать и, главное, принимать решения самостоятельно, без дополнительных управляющих воздействий извне. Конечно, автономная система может принимать и исполнять такие внешние воздействия, но в условиях их отсутствия она вполне может действовать и справляться со своими задачами самостоятельно.

- Адаптивность – это свойство системы, которое позволяет ей действовать в условиях изменчивости внешних воздействий. Если рассматривается техническая система, то адаптивность подразумевает возможность со стороны системы обрабатывать такие входные данные, которые не предусмотрены её изначальным проектом. Например, для нейросетевых систем это свойство обозначает, что система будет адекватно реагировать на такие стимулы, каких не было в обучающей выборке.

Это два главных свойства ИИ-систем, но они не единственные. Давайте изучим ещё три, и в итоге эти пять свойств будут очень точно описывать каждую ИИ-систему, так что их можно будет классифицировать и, соответственно, делать правильные умозаключения.

Третье свойство – то, в какой *реальности* функционирует ИИ-система. Тут имеется два варианта: это либо наша объективная реальность, в которой мы с вами живём, и тогда такая ИИ-система называется «роботом», либо это виртуальная реальность, которая сформирована для ИИ-системы программными методами. Виртуальная реальность может быть связана с нашей, а может быть и не связана. Виртуальный помощник функционирует в виртуальной реальности и воспринимает в ней только голосовые команды. Робот-пылесос работает в нашей реальности и воспринимает её многочисленными сенсорами, а также может воздействовать на неё своими исполнительными устройствами.

Четвёртое свойство – *общность* интеллекта. Здесь тоже имеется дихотомия: искусственный интеллект может быть узким и общим. Узкая ИИ-система решает только одну задачу или один класс задач. Она решает только ту задачу, для которой предназначена. Если ИИ-система играет в шашки, то играть в шахматы она не может, тем более не может водить машину. А вот персональный помощник умеет решать многие задачи, и это первый шаг к построению ИИ-системы общего плана, хотя в этом направлении человечество находится только в самом начале пути.

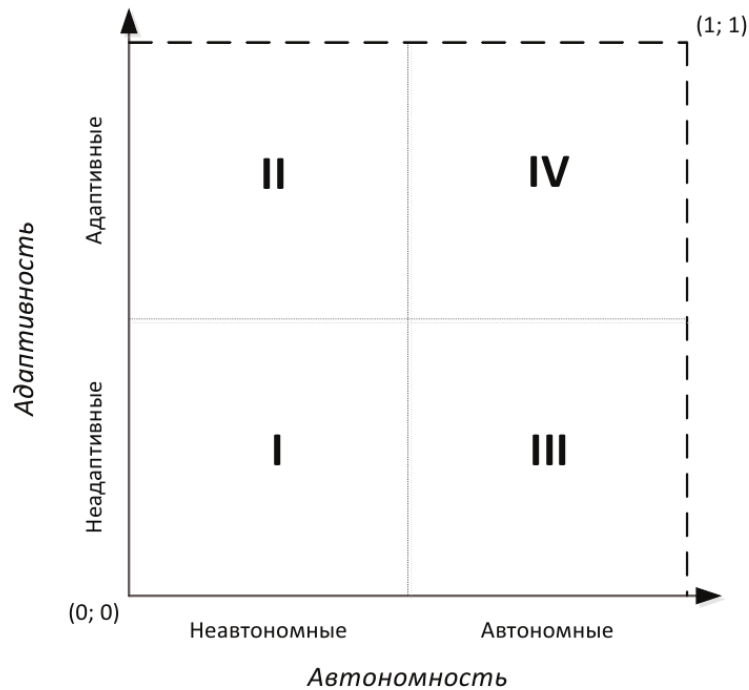
Наконец, пятое свойство. Это тоже дихотомия, которую мы уже рассмотрели в предыдущей главе, – слабый и сильный искусственный интеллект. Напомню для удобства: слабой называется ИИ-система, которая просто решает поставленную задачу, не задумываясь, не рефлексировав, не осознавая того, что делает. Слабый искусственный интеллект – это просто тот самый набор технологий, который позволяет решать нам задачи, относящиеся к прерогативе человека. Все известные нам на сегодняшний день ИИ-системы являются слабыми. А вот сильный искусственный интеллект обладает самосознанием, он осознаёт себя отдельной разумной сущностью. В этом и состоит его отличие. Проблема в том, что мы пока не обладаем необходимыми инструментами, которые позволяют нам ответить на вопрос, обладает ли ИИ-система самосознанием. Мы не можем этого сделать даже относительно другого человека.

И вот теперь имеется пятимерное пространство, разделённое на тридцать два сектора, т. е. по одному сектору на каждый вариант пяти свойств, каждое из которых может принимать два значения. С одной стороны находится неавтономная неадаптивная узкая слабая ИИ-система, работающая в виртуальном мире. С другой стороны – автономный адаптивный общий сильный ИскИн, живущий рядом с нами. Человек, кстати, тоже находится в одном секторе с ИскИном, если только не принимать во внимание естественность его интеллекта. Ну, кстати, шкала «Искусственность – Естественность» является ещё одним измерением, которое, однако, в рамках темы этой книги никакого особого интереса не представляет.

Итак, повторим для закрепления ещё раз. Под интеллектуальностью искусственной (технической) системы понимается наличие у такой системы двух важных свойств. Во-первых, это возможность адаптации к изменяющимся условиям внешней среды при эксплуатации или изменяющимся условиям самой эксплуатации. Во-вторых, это высокая степень автономности её работы, в том числе и в части функциональности по принятию самостоятельных решений. Чем выше степень автономности и адаптивности искусственной системы, тем выше её интеллектуальность.

С другой стороны, под *интеллектуализацией* понимается процесс повышения степени интеллектуальности технической системы. Исходя из того что интеллектуальность является составной характеристикой из двух более простых свойств, интеллектуализация представляет собой траекторию развития системы в рамках своего жизненного цикла от низкого уровня к высокому по двумерному пространству состояний.

Если рассмотреть относительные шкалы для описания и классификации технических систем по степеням их адаптивности и автономности, то можно представить непрерывное двумерное пространство, в котором каждая точка может соответствовать определённой технической системе с заданными степенями адаптивности и автономности от точки (0, 0) – полностью неадаптивная и неавтономная система – до точки (1, 1) – совершенно адаптивная и автономная система. Эта мысль проиллюстрирована следующей схемой.



В качестве типовых примеров систем, которые находятся в квадрантах этого двумерного классификатора интеллектуальности систем, можно привести следующие.

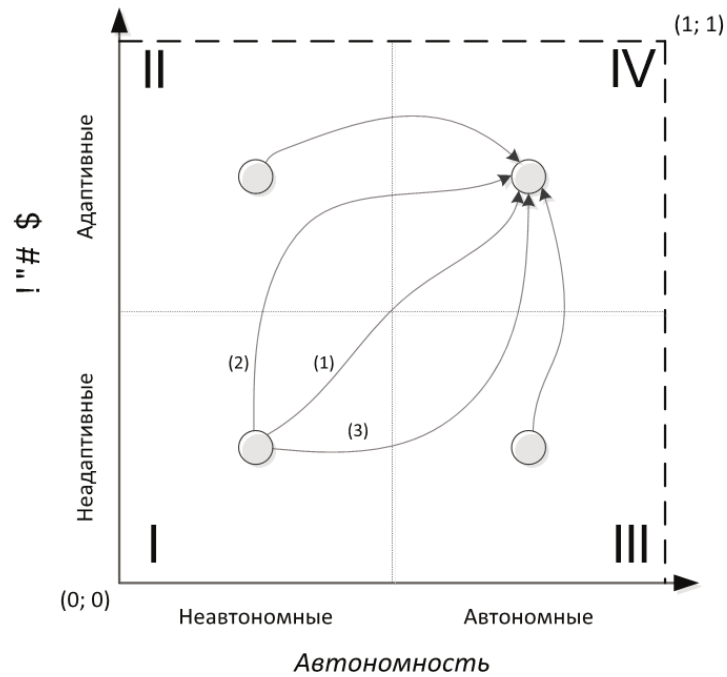
1. *Робот-манипулятор на промышленном производстве*: обычно такой робот предназначен на выполнение заданной последовательности действий (низкая адаптивность) и в принципе не предназначен для продуцирования каких-либо решений (отсутствие автономности).

2. *Робот-пылесос*: этот робот при попадании в незнакомое окружение обследует новое пространство и строит его модель для дальнейшего использования в своей работе (высокая адаптивность), но при этом выполняет только одну функцию уборки помещения с запуском по команде владельца или по триггеру в календаре (низкая автономность).

3. *Система поддержки принятия решений в заданной проблемной области*: такая система обычно уже преднастроена под определённые ситуации, и добавление новой ситуации требует существенной доработки системы (низкая адаптивность), при этом система сама готовит решения и объясняет их (высокая автономность).

4. *Персональный помощник*: система для планирования, напоминания, выполнения рутинных действий, которая постоянно обучается во взаимодействии со своим пользователем (высокая адаптивность) и при этом может самостоятельно запускать различные сценарии общения со своим пользователем и в рамках такого общения самостоятельно делать выбор в пользу того или иного варианта решения на основе прошлых предпочтений пользователя (высокая автономность).

Таким образом, интеллектуализация представляет собой процесс перевода (эволюции) технической системы из её текущего положения в пространстве «Адаптивность – Автономность» в четвёртый квадрант как можно ближе к точке (1; 1). Из квадрантов II и III такой переход может быть осуществлён непосредственно, в то время как в квадранте I может существовать три возможные траектории интеллектуализации системы, что показано на следующей схеме.



Таким образом, предполагается, что процесс интеллектуализации может быть применён к произвольной технической системе. И это довольно важный вывод – при помощи методов и технологий искусственного интеллекта возможно повышение интеллектуальности заданной технической системы посредством изменения её функциональности и внесения в неё двух важных свойств. Ну а делается это при помощи фактически трёх методов, которые рассматриваются далее в этой главе.

Раздел 2.1. Символьные вычисления

Начнём с символьных вычислений. Это, если можно так выразиться, некоторая надстройка над методами представления знаний, которая позволяет представленными знаниями оперировать. Другими словами, символьные вычисления используются для вывода на имеющихся знаниях новых фактов в рамках некоторой проблемной области. Это кардинально отличается от нейросетевого подхода, который будет изучен далее, поскольку каждый шаг символьных вычислений можно объяснить, а сами методы основаны на формальной логике.

Первоначально символьные вычисления основывались на математическом понятии «формальная система», в рамках которого определялся алфавит для построения выражений (он является либо конечным, либо счётным), множество правильно построенных формул (подмножество выражений), множество аксиом (подмножество правильно построенных формул) и множество правил вывода (конечное множество отношений между правильно построенными формулами). Если не вдаваться в подробности, то в рамках формальной системы осуществляется полное абстрагирование от смысла слов естественного языка, на котором выражается какая-то теория, и её перевод в строгие формальные рамки с возможностью при помощи правил вывода синтаксического получения из аксиом и формул других формул. Кроме того, для формальных систем различных типов имеются дополнительные свойства, позволяющие эффективно определять выводимость формул, их разрешимость, непротиворечивость и полноту самой формальной системы.

Обычно множество правил вывода какой-либо формальной системы представляет собой набор посылок с заключениями. Другими словами, каждое правило вывода – это формальное синтаксическое преобразование одной формулы в другую. Такое преобразование может быть записано при помощи нотации «ЕСЛИ... ТО...», а это, в свою очередь, обозначает, что правила вывода формальной системы могут быть описаны продукциями в их самом простом варианте без необходимости использования контекста, условий применимости и других более тонких свойств продукционной модели представления знаний.

От продукций, при помощи которых представляются знания, правила вывода формальной системы ещё отличаются и тем, что выражения в посылках и заключениях таких правил рассматриваются в качестве синтаксических конструкций, а сам вывод представляет собой манипулирование этими конструкциями. Такой подход позволяет реализовать универсальную машину вывода, а процесс интерпретации смысла или семантику вынести на более высокий уровень в рамках интеллектуальной системы. Другими словами, результаты осуществлённого вывода воспринимаются специальным интерпретирующим модулем системы, который посылает команды исполнительным устройствам. Последние уже взаимодействуют с окружением интеллектуальной системы. И для того чтобы она могла называться интеллектуальной, такое взаимодействие должно соответствовать разумному ожиданию акторов, с которыми система взаимодействует.

Необходимо отметить, что некоторые формальные системы содержат счётные бесконечные множества аксиом, правильно построенных формул и даже правил вывода. Это значит, что в явном виде их перечислить невозможно. Вместо явного перечисления используются так называемые схемы – аксиоматические схемы и схемы правил вывода. Так что универсальная машина вывода должна иметь возможность работать и со схемами, на основании которых можно создавать счётные бесконечные множества правил вывода и, соответственно, правильно построенных формул.

Итак, ранее уже были упомянуты логические правила вывода, на которых основывается работа универсальной машины вывода. Во-первых, это правило *Modus Ponens*, которое звучит как «Если есть правило, что из некоторого факта А следует заключение Б, и если при этом

факт А истинен, то можно сделать вывод, что заключение Б тоже истинно». Это правило вывода предназначено для осуществления прямого вывода, когда есть набор фактов (в случае формальной системы – аксиом) и из них необходимо вывести максимальное количество истинных заключений (в формальных системах – правильно построенных формул).

Во-вторых, есть правило *Modus Tollens*, которое звучит как «Если есть правило, что из некоторого факта А следует заключение Б, и при этом установлено, что заключение Б ложно, то можно сделать вывод, что факт А тоже ложен». Это правило вывода используется в обратной стратегии вывода, когда наблюдается некоторое наличное состояние объекта наблюдения в рамках проблемной области и необходимо понять, что могло бы привести к этому состоянию, какие причины лежат в его основе. Обратный вывод обычно даёт вероятностные оценки возможных причин.

Описанные правила вывода основаны на обычной аристотелевой логике. Эти правила использовались в рассуждениях в виде отдельных типов силлогизмов ещё в античном мире. Реализация машины вывода, которая соблюдает такие правила, является делом достаточно простым. Тем не менее это очень мощный формализм, который позволяет решать многие задачи. Однако в конце XX века были предложены новые формализмы, которые расширяют аристотелеву логику и позволяют ещё более тонко подходить ко многим проблемным областям. Рассмотрим некоторые из них.

Первый важный формализм – нечёткая логика и связанные с ней понятия «лингвистическая переменная» и «нечёткая переменная». По мере осуществления попыток формализовать при помощи аристотелевой логики знания о более или менее сложной проблемной области становилось понятно, что простой «чёрно-белый» вариант логики с двумя значениями истинности не может описать всю гамму возможностей при рассмотрении человеком вариантов решения задач. Дело в том, что человек обычно решает задачи не в идеальном мире, а в условиях неполноты информации, нечёткости определений, неточности измерений и т. д. И все такие НЕ-факторы практически невозможно описать при помощи аристотелевой логики. На помощь пришли многозначные логики и, как их апофеоз, бесконечнозначная нечёткая логика. Последняя, к примеру, позволяет осуществлять логический вывод при наличии фактов, не совпадающих с посылками продукций, но машина вывода с возможностью обрабатывать нечёткость всё так же будет способна получить результат, чаще всего вполне приемлемый.

Нечёткий вывод тоже основан на правилах *Modus Ponens* и *Modus Tollens*, но они несколько модифицированы. Первое выглядит как: «Если есть правило, что из некоторого факта А следует заключение Б, и при этом имеется некоторый факт А*, не совсем совпадающий с фактом А, то можно сделать вывод, что заключение Б должно принять вид Б*». В качестве примера можно рассмотреть такой вывод. Пусть есть правило «Если температура окружающей среды низкая, то длительность прогулки короткая». На вход нечёткой машины вывода подаётся факт: «Температура окружающей среды очень низкая», – и при помощи имеющегося правила и этого факта машина вывода делает вывод: «Длительность прогулки очень короткая». Это довольно тривиальный пример, который тем не менее показывает, что нечёткий вывод при помощи небольшого количества правил охватывает огромное число возможностей, в том числе и таких, которые вообще не были ранее предусмотрены разработчиком базы знаний с продукциями. Нечёткий вывод нашёл широкое применение в системах автоматического управления, но он также применяется и в интеллектуальных системах иных классов.

Другим важным классом формальных логик являются так называемые модальные логики. Это довольно обширный класс, и даже простейшая аристотелева логика может быть представлена в виде модальной. Но формализм модальных логик хорош тем, что при помощи него можно обобщать различные понятия и отношения объективной реальности. В частности, такая частная логика, как темпоральная, позволяет формально рассуждать о вопросах, связанных со временем. В этой логике высказывания также имеют временную привязку –

некоторые могут происходить одновременно, другие следовать друг за другом. Другой важной модальной логикой является пространственная логика, которая так же, как и темпоральная, имеет средства выражения пространственных отношений между объектами и высказываниями. Такие высказывания, как «Красный куб находится ближе синего цилиндра», просто формализуются при помощи пространственной логики и очень сложно при помощи аристотелевой. Существуют логики и иных модальностей, но не каждая из них общеприменима, а некоторые эффективны только для каких-либо узких проблемных областей.

Описанные модальные логики имеют важнейшее практическое применение в рамках так называемого пространственно-временного вывода. Чаще всего этот вариант машинного вывода используется в робототехнике, поскольку именно роботы (не только человекоподобные автономные, но и производственные манипуляторы) должны ориентироваться в пространстве и времени. Но также пространственно-временной вывод используется в сложных проблемных областях, в которых требуются рассуждения о том, что было или будет. Чаще всего такие проблемные области относятся к динамическим системам различной природы. В качестве примера можно привести медицину – динамика развития заболевания должна рассматриваться через призму временных отношений, а потому использование временной модальной логики неизбежно.

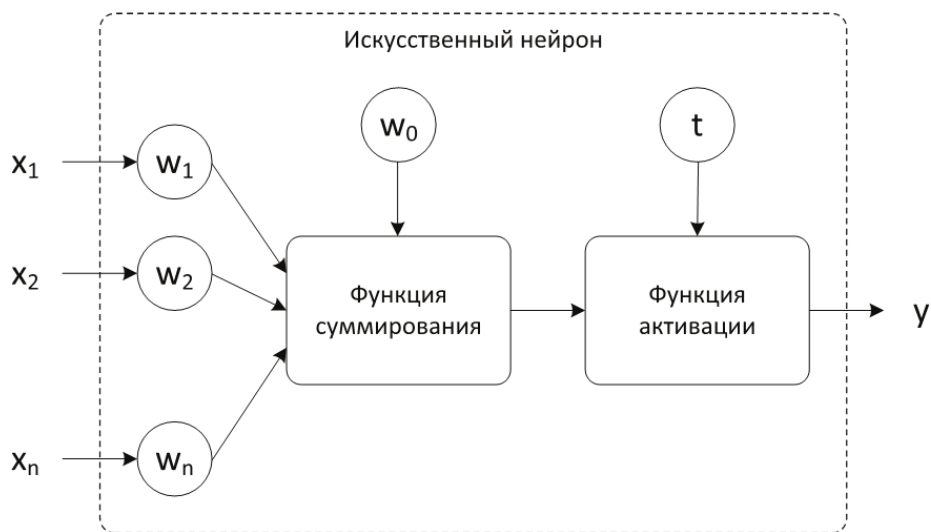
Важнейшим свойством символьных вычислений является возможность объяснить вывод, полученный при помощи них. Если есть начальные факты, есть последовательность рассуждений и есть множество правил вывода, то можно объяснить, почему из начальных фактов были получены такие-то результаты. Это очень важное свойство интеллектуальных систем, и оно в большей мере соответствует нисходящему подходу в построении искусственного интеллекта.

Раздел 2.2. Искусственные нейронные сети

Искусственная нейронная сеть – это в первую очередь математическая модель машинного обучения, которая решает задачи примерно по тем же принципам, что и биологические нейронные сети, состоящие из нервных клеток. Искусственные нейронные сети имеют программную или аппаратную реализацию, поэтому чисто абстрактная математическая модель воплощена в действительности в виде вычислительной системы. В то же самое время искусственные нейронные сети представляют собой самую широко распространённую модель машинного обучения и одновременно с этим являются типичным подходом в рамках восходящего искусственного интеллекта.

Искусственная нейросеть состоит из множества взаимодействующих друг с другом искусственных нейронов, чаще всего собранных в слои так, что нейроны одного слоя получают информацию только от нейронов предыдущего слоя и передают информацию на следующий слой. Сам по себе искусственный нейрон представляет собой *очень* упрощённую модель биологического нейрона. У искусственного нейрона есть множество входов, один выход, а также две функции: суммирования и активации. Впрочем, существуют и другие типы искусственных нейронов, в том числе и таких, работа которых зависит от времени. Фактически каждый искусственный нейрон решает задачу простой логистической регрессии (в случае если его функция активации представляет собой сигмоиду).

Первую модель искусственного нейрона предложили Мак-Каллок и Питтс в 1943 году. Эта модель выглядит следующим образом.



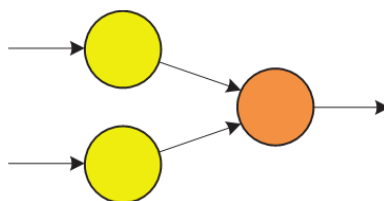
Чаще всего в качестве упрощения считается, что входные переменные x_1, x_2, \dots, x_n и выходное значение y принимают значения из интервала $[0, 1]$. Функция суммирования представляет собой обычную взвешенную сумму, результат которой равен сумме произведений каждого входного значения x_i на свой вес w_i , а некоторый специальный вес w_0 добавляется всегда. Наибольший интерес представляет функция активации, которая может принимать различные формы, так что свойства как искусственного нейрона, так и всей нейросети часто зависят от выбора формы функции активации.

Результат функции суммирования передаётся на вход функции активации. В свою очередь, её результат является выходом нейрона y , который передаётся на вход следующим нейронам. Обычно, как уже было указано, функцию активации подбирают так, чтобы область её

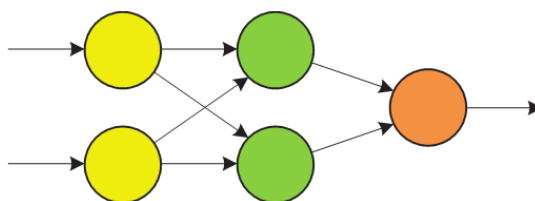
значений лежала в интервале $[0, 1]$, хотя это и необязательно. В первых реализациях искусственных нейронных сетей функция активации вообще была дискретной с областью определения $\{0, 1\}$, и такие нейросети неплохо справлялись с задачами классификации. Сегодня в сетях глубокого обучения могут использоваться разные функции активации у разных нейронов, но обычно все они входят в класс сигмоид – это и логистическая функция, и гиперболический тангенс, и функция Гаусса, и многие другие. Выбор функции активации зависит от многих параметров задачи, и часто архитектор, проектирующий искусственную нейросеть, должен проявить для этого немалый уровень творчества.

Со времён, когда первые разработчики искусственных нейронных сетей сделали свои первые искусственные нейросети, были созданы многочисленные модификации модели, которые подходят для решения более специфичных задач. Конечно, для погружения в теорию искусственных нейросетей необходимо читать специализированную литературу и изучать курсы, но здесь можно кратко охарактеризовать многие интересные модели и решаемые ими задачи. Так что начнём...

Несомненно, первой реализованной моделью искусственной нейронной сети был перцептрон Ф. Розенблатта, который тот предложил в 1957 году. Эта модель была реализована «в железе», первым в мире «нейрокомпьютером» стал компьютер Марк-1, построенный под руководством Розенблатта в 1960 году. Перцептрон – это простая нейронная сеть с тремя слоями: входным, скрытым и выходным. Таким образом, перцептрон реализует простейшую кибернетическую машину с сенсорами (слой входных нейронов), управляющим устройством (слой скрытых нейронов) и эффекторами (слой выходных нейронов). В перцептроне используется пороговая передаточная функция и прямое распространение сигнала. Как математическая модель перцептрон уже был достаточно мощным формализмом для решения большого количества задач, поскольку он на основе обучения позволял классифицировать, кластеризовать и прогнозировать, т. е. решать большинство классических задач машинного обучения. Впрочем, после публикации книги М. Минского и С. Паперта «Перцептроны», в которой авторы показали принципиальную невозможность для перцептрона решить некоторые задачи (сюрприз – задача «XOR», традиционно включаемая в класс нерешаемых перцептроном, на самом деле к таковым не относится), постепенно интерес к перцептрону снизился, и большее внимание стала получать нисходящая парадигма в искусственном интеллекте, при этом сам Марвин Минский был её оппонентом. Тем не менее сегодня с развитием математического аппарата и средств вычислительной техники интерес к перцептрону и его расширениям вновь вырос.

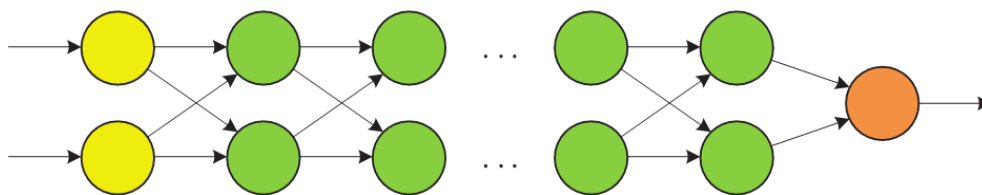


Однослойный перцептрон



Классический перцептрон

Необходимо отметить, что в процессе развития перцептронов появились некоторые расширения первоначальной модели, предложенной Ф. Розенблаттом. Самый простой классификатор основан на подсчёте слоёв в перцептроне: однослойный, с одним скрытым слоем (классический) и многослойный. Все эти типы были в своё время описаны Розенблаттом. Другая авторская классификация включала: элементарный перцептрон, простой перцептрон, перцептрон с последовательными связями, перцептрон с перекрёстными связями, перцептрон с обратными связями, перцептрон с переменными связями. Первые три класса были описаны самим автором, а следующие три развиты в дальнейшем при детальной проработке модели искусственных нейронных сетей.



Многослойный перцептрон

Из-за первоначальной неразберихи в терминологии и повышенных ожиданий, которые появились в отношении перцептрона и модели искусственного нейрона, Дэвидом Румельхартом был предложен новый класс перцептронов, которые сейчас называются «многослойными перцептронами Румельхарта» и отличаются от многослойных перцептронов Розенблатта тем, что для обучения в них используется метод обратного распространения ошибки, в то время как у Розенблатта использовался метод коррекции ошибки. Есть ещё несколько отличий, в частности у Румельхарта в качестве функции активации используется сигмоида, а число обучаемых слоёв больше одного.

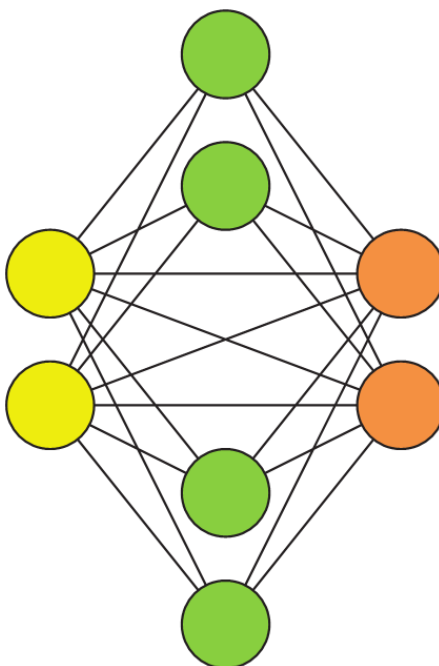
Фактически перцептрон – это самый простой пример нейронной сети прямого распространения. Другим интересным примером является сеть ELM, экстремальная обучающаяся машина. В этой нейросети нейроны не располагаются в слое, а связаны друг с другом случайным образом. Выделяются только входной и выходной слои, а остальные нейроны находятся между ними и связаны друг с другом именно случайно. Обучение сети также производится методом обратного распространения ошибки. А вот если обучение производится при помощи обновления состояния нейронов по результатам наблюдения за их работой, за порядком активации, то это уже нейронная эхо-сеть, ESN.

Далее некоторое время модель искусственных нейронных сетей развивалась в сторону тонкого подбора функций активации (например, в сплайн-модели Хакимова функция активации реализована в виде сигмоидального сплайна, либо в сети Брумхеда-Лоу используются радиально-базисные функции), типов весовых коэффициентов между нейронами и других параметров, не влияющих на структуру сетей. Это не позволяло выходить на новые объёмы данных или решать новые задачи, но было обусловлено скорее тем, что на существующих в те времена вычислительных мощностях было затруднительно реализовывать нейронные сети больших размеров.

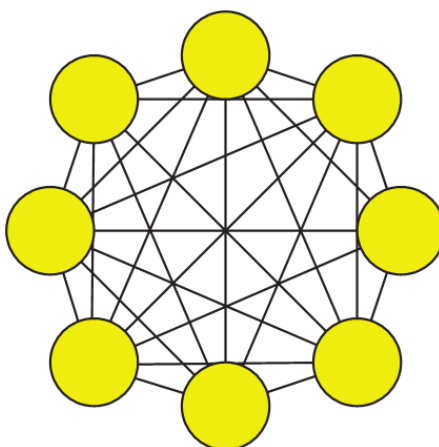
Однако перцептрон получал развитие и с точки зрения изменения структуры. Так, например, Джефф Элман и Майкл Джордан предложили свои варианты изменения перцептрона и включения в него обратных связей так, чтобы получилась рекуррентная сеть. В варианте Джордана на вход многослойного перцептрона также подаются выходные значения с задержкой от одного и более тактов. В нейронной сети Элмана используется такой же подход, только рекуррентные связи с задержкой в несколько тактов подаются не с выходных нейронов, а со скрытых слоёв многослойного перцептрона. Оба варианта структурного развития перцептрона обла-

дают определённой памятью. Впрочем, различными вариантами памяти обладают все рекуррентные сети.

Ещё один интересный вариант нейросети с небольшим количеством скрытых слоёв – сеть Ворда, в которой имеется только один внутренний слой, но его нейроны разбиты на группы. В каждой группе используется своя передаточная функция, и результаты работы каждого блока передаются на выходной слой, который как бы рассматривает вход с разных точек зрения. Кроме того, входной слой может напрямую замыкаться на выходной, и это тоже позволяет добавить гибкости сети. Топология конкретной реализации сети Ворда определяется количеством блоков в скрытом слое и наличием прямого замыкания входных нейронов на выходные. Сеть показывает отличные результаты для задачи распознавания образов.



Сеть Ворда

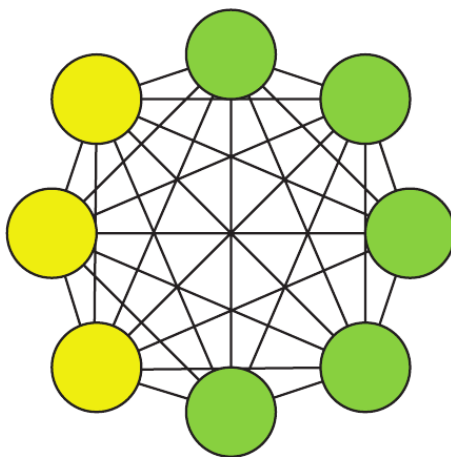


Сеть Хопфилда

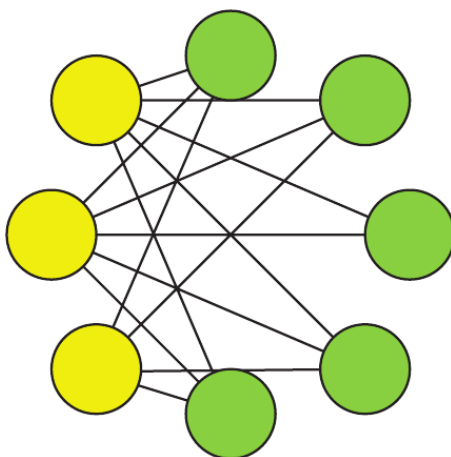
В 1982 году Джон Хопфилд предложил интересный вариант однослойной нейросети, который позволял достаточно просто решать некоторые задачи оптимизации или предоставлять механизм автоассоциативной памяти. В этой сети имеются только входные нейроны, каждый из которых соединён со всеми другими (т. е. сеть представляет собой полносвязный граф).

Обучение сети также своеобразно – для всех связей между нейронами необходимо аналитически рассчитать веса, и это значит, что обучение длится один цикл. Обучение (или аналитический расчёт весов) как бы сохраняет в весовых коэффициентах сети набор эталонных образов, которые сеть должна запомнить. После того как сеть обучена, её можно запускать в рабочем режиме, и он также отличается от традиционного единичного прогона. Сеть Хопфилда работает до тех пор, пока её следующее состояние не будет равно предыдущему, т. е. сеть должна достигнуть равновесия. Достигнутое равновесное состояние обязательно соответствует одному из запомненных эталонных образов. Тем самым осуществляется распознавание и даже восстановление повреждённых образов.

Развитием сети Хопфилда стала так называемая неограниченная машина Больцмана, которая представляет собой стохастический вариант этой сети. Машина Больцмана представляет собой рекуррентную нейронную сеть, которая для обучения использует алгоритм имитации отжига. Неограниченная машина позволяет решать сложные комбинаторные задачи, однако её практическое применение затруднительно или даже невозможно из-за комбинаторного взрыва в части гигантского количества связей между нейронами. Однако если использовать ограничение на количество связей между нейронами, то машина Больцмана становится менее мощной, но её использование вполне возможно. В частности, из каскадов из машин Больцмана составляются глубокие сети доверия – сначала сеть обучается при помощи стандартного алгоритма для машин Больцмана, а потом дообучается при помощи обратного распространения ошибки.

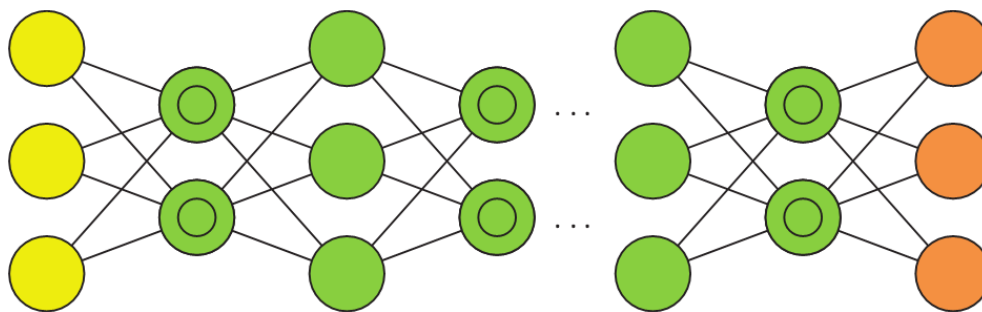


Машина Больцмана



Ограниченная машина Больцмана

Итак, рассмотрим каскадное объединение ограниченных машин Больцмана, при котором скрытые нейроны одной машины становятся входными для следующей. Именно так формируется глубокая сеть доверия. Такая сеть также обучается без учителя, причём обучение происходит послойно. Как и в случае сети Хоп-филда и машины Больцмана, сеть глубокого доверия осуществляет кластеризацию и восстановление входов, а при дополнительном обучении с учителем осуществляется и классификация, которая может быть очень точной. Именно на глубоких сетях доверия были построены первые алгоритмы глубинного обучения.



Глубокая сеть доверия

Ещё один пример глубокой нейронной сети прямого распространения – это глубокая остаточная сеть (DRN). В ней количество скрытых слоёв может быть очень велико, и прямые связи от слоя передаются не только на следующий слой, но и на дальнейшие слои. При этом на каждый слой также передаётся входная информация в неизменённом виде, так что каждый слой получает как бы пару (входная информация, выходная информация предыдущего слоя). Такая архитектура может очень неплохо распознавать образы, однако было показано, что подобная нейросеть тождественна рекуррентным нейросетям без учёта времени.

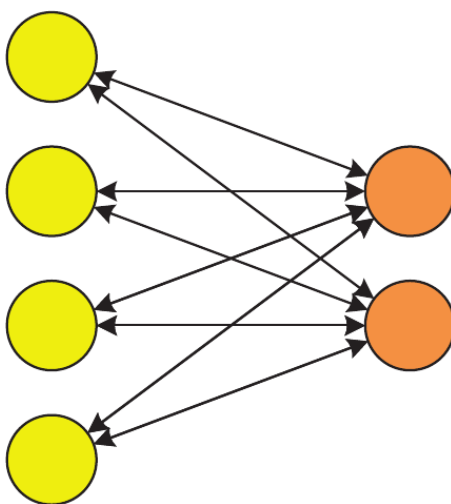
Так что теперь более подробно рассмотрим эти самые рекуррентные нейронные сети, т. е. такие, в которых возможно либо замыкание выхода нейрона на его же вход, либо создание обратных связей от нейронов дальних слоёв к нейронам ближних. При этом значение на рекуррентные связи подаётся с определённой временной задержкой, измеряемой в тактах работы нейронной сети. В принципе, и сеть Хопфилда, и машина Больцмана уже сами по себе являются рекуррентными сетями, но они представляют самый примитивный их класс. Теперь пришло время познакомиться с более продвинутыми.

Одна из самых первых рекуррентных нейронных сетей была предложена в 1987 г. – так называемая сеть Хэмминга, которая является расширением сети Хопфилда и решает задачу классификации бинарных векторов, критерием различия которых является расстояние Хэмминга. Сеть Хэмминга представляет собой трёхслойную сеть, при этом количество нейронов во втором и третьем слоях равно количеству классов классификации, и со второго и третьего слоёв есть рекуррентные связи – со второго на первый, а с третьего на третий же. После обучения сеть выдаёт номер класса, которому принадлежит образец, поданный на вход. Эта сеть очень эффективно работает на довольно узком классе задач. В частности, она может использоваться для распознавания чёрно-белых изображений: штрих-кодов, цифр индекса и др.

В 1988 г. Барт Коско разработал тип нейронной сети, названный им «двунаправленной ассоциативной памятью» (а сегодня она называется сетью Коско). Это вариант сети Хопфилда, в котором входными является часть нейронов, а выходными – оставшиеся нейроны. Все нейроны входного и выходного слоёв соединены друг с другом двунаправленными связями, которые обучаются так же, как и для сети Хопфилда. Сеть Коско работает на тех же задачах, что и сеть Хопфилда, но делает это немногим более эффективно. Также разработаны некоторые

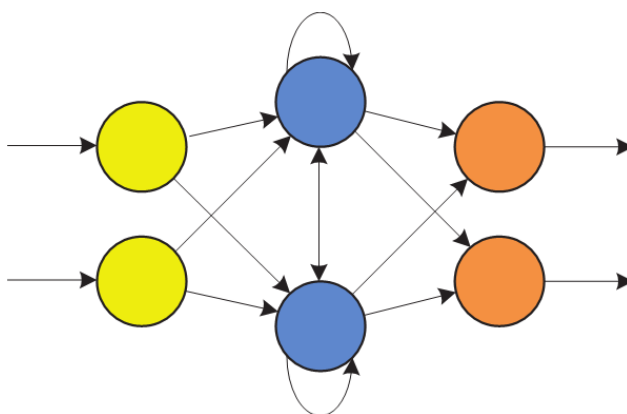
подтипы этой сети, в частности адаптивный вариант, в котором осуществляется изменение весов связей в процессе работы. Тем самым адаптивная сеть Коско позволяет идентифицировать искажённые и зашумлённые образцы.

Вообще, рекуррентными нейронными сетями называют любые искусственные нейросети, в которых есть обратные связи от нейронов более глубоких слоёв к нейронам менее глубоких, либо замыкание нейронов самих на себя или на нейроны того же самого слоя. При этом такая архитектура связей между нейронами и слоями может создаваться и в уже рассмотренных типах сетей, в которых изначально рекуррентных связей не предполагалось. Подобные сети могут запоминать информацию, организовывать циклы и иметь другие интересные свойства. Однако из-за сложности анализа поведения таких моделей большая часть предложенных или потенциальных архитектур изучена плохо.



Сеть Коско

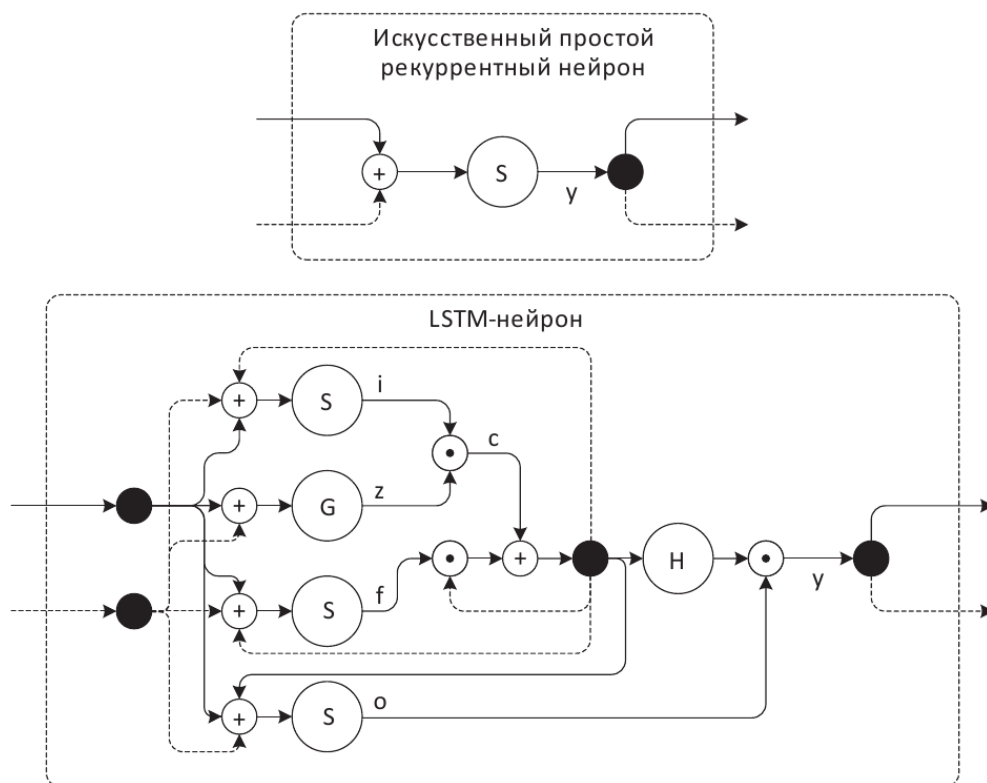
Общий вид архитектуры рекуррентной нейронной сети можно представить следующим образом.



В процессе исследований рекуррентных нейронных сетей стало ясно, что с обычными искусственными нейронами они работают не совсем эффективно – обычные нейроны с течением времени как будто бы «размываются», теряют своё обучение. Это не позволяет эффективно использовать обычные нейроны в рекуррентных нейронных сетях для задач классификации и, главное, прогнозирования временных рядов. И тогда появилась идея по разработке

новых типов искусственных нейронов. Одним из наиболее знаменитых в этом плане является LSTM-нейрон (от *англ.* LongShort-TermMemory – «долгая краткосрочная память»).

Итак, Сепп Хохрайтер и Юрген Шмидхубер в 1997 г. предложили новый вариант искусственного нейрона, который затем был улучшен Феликсом Герсом в 2000 г. Искусственная нейросеть, составленная из LSTM-нейронов, является универсальной в смысле Тьюринга, т. е. при помощи неё можно вычислить любую вычислимую функцию (матрицу весов такой нейросети можно воспринимать как программу). Именно такие нейросети сегодня являются фундаментальной основой для построения решений ведущих компаний, работающих в области искусственного интеллекта. Сам же LSTM-нейрон может быть представлен при помощи следующей диаграммы.



Представленные выше диаграммы нарисованы с использованием следующих условных обозначений:

- обычная стрелка – входные и выходные связи нейрона, поток значений между функциями внутри нейрона;
- пунктирная стрелка – входные и выходные связи нейрона, поток значений между функциями внутри нейрона с задержкой по времени (на один или более тактов работы нейронной сети);
- плюс – функция суммирования входных значений;
- точка – функция перемножения входных значений;
- чёрный круг – дубликатор потока значений;
- гейт **S** – функция активации, обычно представляющая собой сигмоиду;
- гейт **G** – функция активации для входного потока, обычно являющаяся гиперболическим тангенсом;
- гейт **H** – функция активации для выходного потока, обычно тоже гиперболический тангенс;
- поток **i** – результат активации для входного потока;

- поток z – результат активации для рекуррентного входного потока (обычно с нейронов того же слоя);
- поток f – результат активации для гейта забывания;
- поток o – результат активации для выходного потока;
- поток y – результат работы LSTM-нейрона в целом.

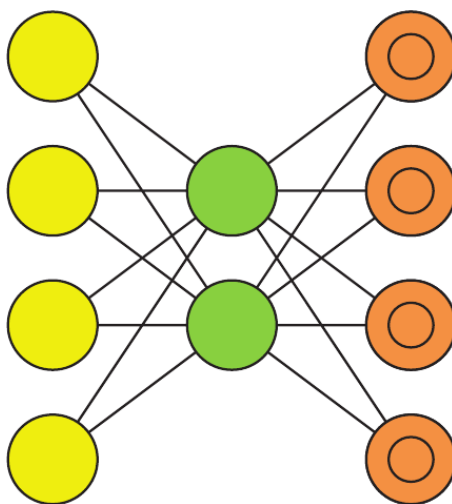
На сегодняшний день, как уже сказано выше, LSTM-нейрон является самой продвинутой моделью биологического нейрона. Однако он всё так же имеет одно важное упущение – нейроны в нервной системе млекопитающих получают информацию не только при помощи передачи электрических импульсов от дендритов через аксоны к следующим нейронам, но ещё используются различные нейромедиаторы в синаптических щелях, а также гуморальное воздействие на весь нейрон в целом. Кроме того, до сих пор не до конца изучена роль глиальных клеток, которых в нервной системе на порядок больше, чем нейронов. Другая сложность, которая проявляется при реализации искусственных нейросетей на LSTM-нейронах, – их относительно высокие требования к вычислительным ресурсам.

Из-за описанных сложностей с LSTM-нейроном был разработан так называемый GRU-нейрон (от *англ.* *gated recurrent unit* – «рекуррентный нейрон с гейтами»), который практически повторяет структуру LSTM-нейрона, но в нём нет выходного гейта **H** и, соответственно, всех используемых этим гейтом потоков. Это позволило существенно снизить вычислительную сложность работы такого искусственного нейрона, однако он так же эффективен для отдельных задач, как и LSTM-нейрон (в частности, на GRU-нейронах хорошо решаются задачи синтеза – моделирование музыки или моделирование текста и разговора).

Мы изучили сети прямого распространения и рекуррентные сети, и уже по пройденному материалу видно, как широко это направление исследований. Так что теперь мы можем перейти к более прикладным вариантам, на которых сегодня решается большинство задач, где используются искусственные нейронные сети.

Самый первый интересный вариант – так называемый «автокодировщик». Этим словом обозначается специальная архитектура искусственной нейронной сети с одним скрытым слоем без обратных связей. Главная особенность такой сети заключается в том, что количество нейронов на входном и выходном слоях одинаково, а на скрытом слое – меньше. Это ограничение необходимо для того, чтобы сеть не зафиксировалась на тривиальном варианте в процессе обучения без учителя, которое для автокодировщиков заключается в получении на выходном слое результатов, как можно более похожих на входные параметры сети. Другими словами, обучение нейронов скрытого слоя в таких условиях позволяет автокодировщику как бы осуществлять обобщение, сжимающее кодирование входной информации, отсюда и такое наименование.

Автокодировщики часто используются для последовательного обучения слоёв искусственных нейронных сетей. При таком подходе каждый новый слой представляется в качестве скрытого слоя автокодировщика, при этом входным слоем является предыдущий слой, обученный на предыдущем же шаге. Размерность каждого следующего слоя снижается, и тем самым нейросеть должна находить всё больше и больше обобщений, она должна всё сильнее и сильнее повышать степень абстракции кодируемой информации. После такого обучения нейросеть на каждом скрытом слое содержит нейроны, которые отвечают за очень глубокие уровни абстракции входной информации, и чем глубже слой, на котором находится нейрон, тем выше уровень абстракции. Сегодня это один из основных подходов в технологии глубокого обучения.



Автокодировщик

Перед переходом к рассмотрению глубоких нейронных сетей остаётся отметить несколько особенных вариантов автокодировщиков.

- *Вариационный автокодировщик* – автокодировщик, в рамках которого делаются сильные предположения относительно распределения скрытых переменных. Такие автокодировщики используют вариационное исчисление для обучения представлению скрытых переменных, что приводит к дополнительной потере компонентов при использовании специализированного алгоритма обучения. Он предполагает, что данные сгенерированы направленной графической моделью, и автокодировщик при обучении строит аппроксимацию апостериорного распределения.

- *Автокодировщик, удаляющий шум* – автокодировщик, который принимает частично «повреждённые» входные данные, а обучение осуществляется для получения неискажённого выхода. Такие автокодировщики восстанавливают так называемое «хорошее представление», которое может быть получено из повреждённого выхода так, что достаточно точно воспроизводит ожидаемый выход.

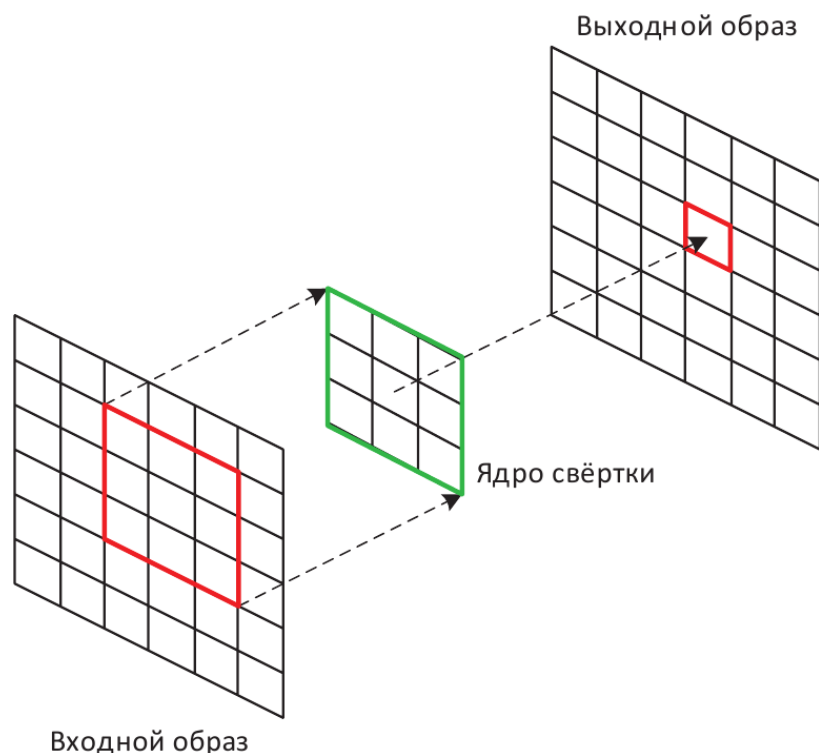
- *Разрежённый автокодировщик* – автокодировщик, который во время обучения имеет большее число скрытых нейронов, чем входных. Это создаёт так называемую «разрежённость» в нейронах, которая накладывается и на выходные нейроны. Иногда это бывает полезно при решении задач классификации.
- *Контрастный автокодировщик* – автокодировщик, который явным образом добавляет регуляризацию своей целевой функции, что заставляет нейронную модель исследовать функцию, которая является устойчивой к небольшим изменениям входных значений.

Теперь, наконец-то, перейдём к рассмотрению искусственных нейронных сетей, которые получили название «свёрточных» (а также двойственных им – «развёрточных»). Такие нейросети получили в последнее время самое широкое распространение, поскольку являются базовым строительным блоком для построения систем «глубокого обучения». Рассмотрим несколько примеров.

Свёрточная сеть – ещё одна специальная архитектура искусственных нейронных сетей, которая очень эффективно решает задачу распознавания образов. Сама по себе архитектура была предложена в 1988 г. французским учёным Яном Лекуном, который взял некоторые идеи из устройства нервной системы млекопитающих и, в частности, зрительной коры головного мозга. Структура свёрточной нейронной сети состоит из большого количества слоёв двух чередующихся типов – свёрточных и субдискретизирующих. Обычно работа такой сети описывается как постепенный переход от конкретных особенностей распознаваемого образа к

абстрактным деталям всё выше и выше по иерархии уровней абстракции. Например, в применении к изображениям это обозначает, что первые слои обнаруживают в частях изображения различные типовые элементы вроде разнонаправленных отрезков, а на последних слоях нейроны уже активируются в ответ на наличие в составе изображения каких-либо форм, конкретных объектов или образов.

Суть операции свёртки, из-за которой сеть получила своё название, заключается в том, что каждый фрагмент входного образа поэлементно умножается на матрицу свёртки, результат этого перемножения суммируется, и итоговая сумма передаётся в аналогичную позицию выходного образа для заданного слоя.



Каждый слой свёртки передаёт свой выход на слой субдискретизации, и это важная особенность рассматриваемой архитектуры. Субдискретизация нелинейно уплотняет карту признаков так, что группа пикселей (обычно 3×3) преобразуется в один пиксель, и это делается при помощи нелинейного преобразования (обычно используется функция максимизации). Суть этой операции в том, чтобы получить карту признаков с предыдущего слоя свёртки и «огрубить» изображение. Другими словами, происходит отказ от ненужных деталей, и нейронная сеть начинает рассматривать всё более и более абстрактные признаки исходного изображения. Кроме максимизации в качестве операции субдискретизации, могут использоваться и другие, и от этого зависят свойства получаемой нейронной сети, лучше или хуже подходящие для решения поставленной задачи.

Наибольший интерес вызывает то, что в операции свёртки матрица или ядро свёртки изначально не закладывается разработчиком нейронной сети, а самостоятельно подбирается в процессе обучения нейронной сети (чаще всего методом обратного распространения ошибки). В итоге для каждого свёрточного слоя получается большое количество карт признаков – отрезки и дуги, направленные под разными углами, границы между сплошными средами, точки и иные примитивы; а на более глубоких слоях – лица, животные, автомобили, здания и т. д. Количество карт признаков для каждого слоя является метапараметром сети и определяется разработчиком.

Для чего нужна свёрточная нейронная сеть? Как показали эксперименты, свёрточные нейросети:

- являются одним из лучших методов по распознаванию образов;
- по сравнению с полносвязными архитектурами (например, с перцептронами), требуют намного меньшего количества вычисляемых в процессе обучения параметров, из чего следует возможность построения сетей с большим количеством слоёв;
- позволяют эффективно параллелизовать свои вычисления;
- обладают устойчивостью к поворотам и сдвигам распознаваемых образов.

Если про искусственные нейронные сети можно написать отдельную книгу, то про свёрточные нейронные сети можно написать ещё одну, поэтому оставим их и перейдём к следующему типу архитектуры – развёрточным сетям.

Развёрточные нейронные сети – это как бы перевернутые задом наперёд свёрточные сети. Они были предложены Мэтью Зайлером для простой задачи – анализа работы свёрточных нейронных сетей. Дело в том, что глубинное обучение задействует воистину огромное количество слоёв, а в свёрточных нейросетях эти слои ещё и разных типов, и субдискретизация осуществляется по разным каналам, а с учётом того, что чем глубже и полносвязнее сеть, тем сложнее понять, как она обучается, необходим хотя бы какой-то эвристический механизм или инструмент для оценки того, правильно ли обучается свёрточная нейросеть и не переобучилась ли она.

Если говорить самым простым языком, то развёрточная нейронная сеть строит иерархические представления свёрточной нейросети, к которой подключена. При этом в рассмотрение принимаются все карты признаков и параметры, которые были получены во время обучения свёрточной нейросети. Развёрточная нейронная сеть как бы пытается восстановить тот сигнал, который распознавала свёрточная нейросеть, но из-за большого количества нелинейных и необратимых преобразований это удаётся сделать лишь частично. Однако развёрточная нейросеть подключена к каждому свёрточному слою свёрточной нейросети и восстанавливает изображения для всех свёрточных слоёв, обучаясь параллельно. В итоге получается нейронная сеть, которая позволяет «видеть» то, как обучена свёрточная нейронная сеть и хотя бы отчасти объяснять результаты.

Перейдём к следующему важному варианту архитектуры искусственных нейронных сетей. Это генеративно-сопоставительная сеть (GAN), которая предназначена для обучения без учителя. Генеративно-сопоставительная нейронная сеть составлена из двух подсетей. Одна из них (сеть G – генеративная сеть) генерирует различные образцы, а вторая (сеть D – дискриминационная сеть) – отличает правильные образцы от неправильных. Для целей машинного обучения сети G и D имеют противоположные задачи, и между ними организуется «антагонистическая игра», т. е. игра, в которой участвуют два игрока, выигрыши которых противоположны. Сеть G генерирует образец (например, изображение), начиная со смешивания некоторых исходных образцов, используя случайные значения своих скрытых параметров (весовых коэффициентов). После генерации образца сеть D пытается отбраковать те образцы, которые выглядят неприемлемо, и результат работы этой сети подаётся на вход сети G, которая при помощи метода обратного распространения ошибки перестраивает свои весовые коэффициенты так, чтобы «обмануть» сеть D. В итоге на каждой итерации генерируемые образцы становятся всё лучше и лучше.

Обычно в качестве сети D берётся свёрточная сеть глубокого обучения – ну просто потому, что именно свёрточные сети, как уже описано, лучше всего занимаются распознаванием образов, а дискриминационная сеть должна именно распознавать образы, которые создаёт генеративная сеть. А в качестве последней используются те сети, которые могут генерировать образы. И вообще говоря, сегодня именно генеративные нейронные сети находятся на пике интереса, и здесь могут быть самые разные варианты нейросетей. Например, генера-

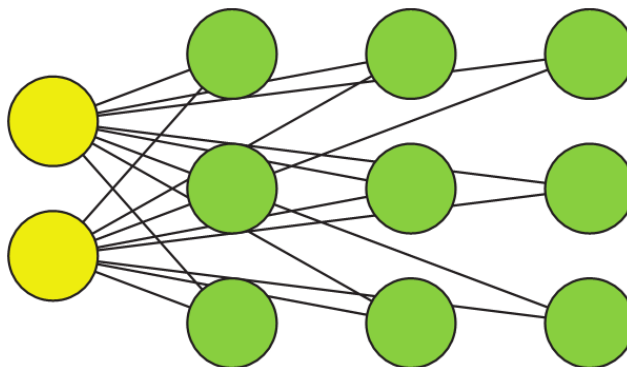
тивно-состязательные нейронные сети нашли очень широкое применение в вопросе так называемого машинного творчества.

Финский учёный в области искусственного интеллекта Теуво Кохонен предложил несколько интереснейших концепций в рамках машинного обучения и искусственных нейронных сетей, из-за чего он стал наиболее часто цитируемым финским учёным. В частности, им предложены:

- фундаментальная теория ассоциативной памяти;
- особенный алгоритм обучения нейронных сетей;
- особый класс нейронных сетей (нейронные сети Кохонена);
- самоорганизующиеся карты Кохонена;
- модель нейрона и специального слоя (слой Кохонена).

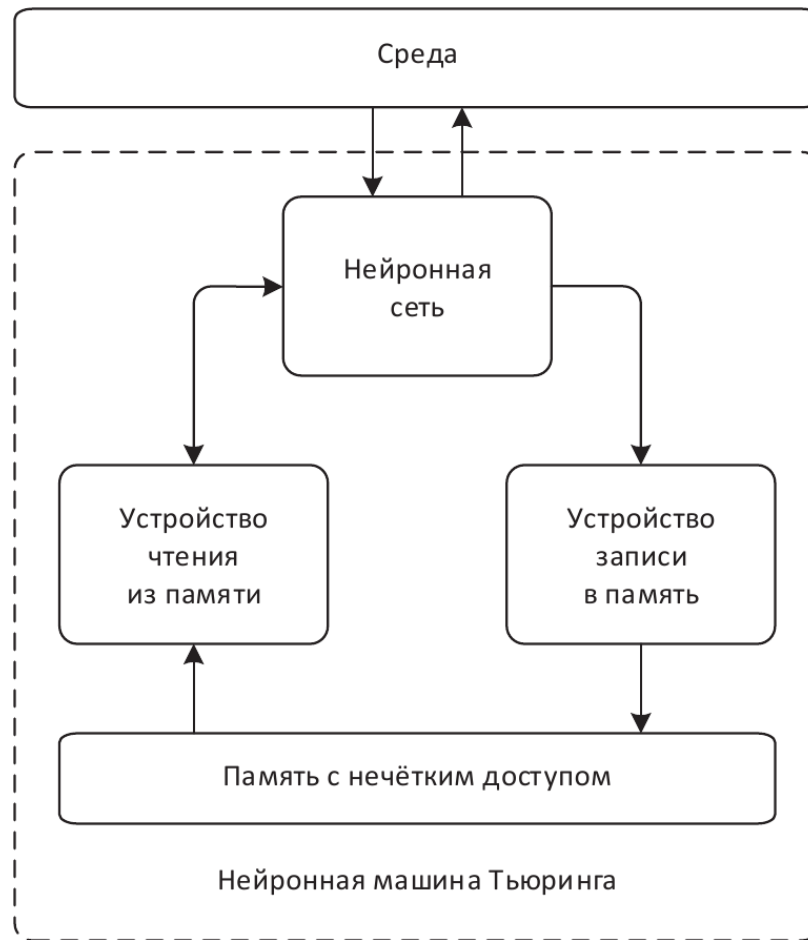
Наибольший интерес представляет самоорганизующаяся карта Кохонена. Это особая архитектура нейронной сети Кохонена для обучения без учителя. Карта решает задачи кластеризации и снижения размерности и применяется для решения задач моделирования, прогнозирования, выявления наборов независимых признаков, поиска закономерностей в больших массивах данных, квантизации признаков к их ограниченному числу индексов и некоторых других.

Самоорганизующаяся карта Кохонена получает на вход массив многомерных данных и проецирует его на двумерную плоскость, «раскрашивая» её и получая что-то вроде разноцветной административной карты территории (отсюда и название архитектуры). От разработчика или аналитика необходимо только задание количества кластеров, на которые необходимо разбить входной массив, а дальше нейросеть всё сделает самостоятельно.



Самоорганизующаяся карта Кохонена

Очень интересный подход предложили несколько учёных в 2014 г. В рамках него искусственную нейронную сеть совместили с внешней памятью, ведь произвольный доступ к памяти является критически важной функцией при обработке информацией. И компьютеры, и мозг человека работают с тем или иным видом памяти, к которой можно обращаться более или менее произвольно. Так вот, в предлагаемой архитектуре, которая была названа «нейронной машиной Тьюринга», нейронная сеть получает входную информацию и выдаёт выходную информацию не только во взаимодействии с внешним миром, но и использует внутреннюю память в виде матрицы чисел. Ячейки этой матрицы-памяти индексируются при помощи так называемых нечётких индексов, что позволяет нейросети обращаться ко всем ячейкам как бы одновременно, получая их линейную комбинацию.



На диаграмме показана общая структурная схема нейронной машины Тьюринга. Нейронная сеть является управляющим устройством машины и получает из среды входную информацию, выдавая в неё выходную информацию. Нейронная сеть управляет устройствами чтения и записи в память, которые, соответственно, осуществляют чтение и запись в память с нечётким доступом. Обращение к памяти осуществляется на каждом цикле работы нейронной сети, но при этом она сама обучается, как и когда это делать.

Эта архитектура, как предполагается, имеет много интересных приложений. Уже сегодня в экспериментах по распознаванию образов, обработке текста и по решению некоторых других задач такая нейронная машина Тьюринга показала результаты, превосходящие по производительности рекурсивные сети с LSTM-нейронами.

В большинстве моделей для кластеризации данных предполагается, что количество кластеров определяется заранее до старта алгоритма, сам алгоритм кластеризации работает один раз, а при изменении входных данных его необходимо запускать повторно (возможно, с изменением числа кластеров). Это схема работы большинства моделей машинного обучения, в том числе и многих нейросетевых моделей. В качестве альтернативной модели была предложена модель расширяющегося нейронного газа, в которой нет этих недостатков, – в ней число кластеров определяется самостоятельно, а сама модель является динамической, так что при изменении входных данных она тут же пересчитывает свои параметры и выдаёт новый результат, основанный на истории своей работы, в том числе и до изменения входных данных.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.