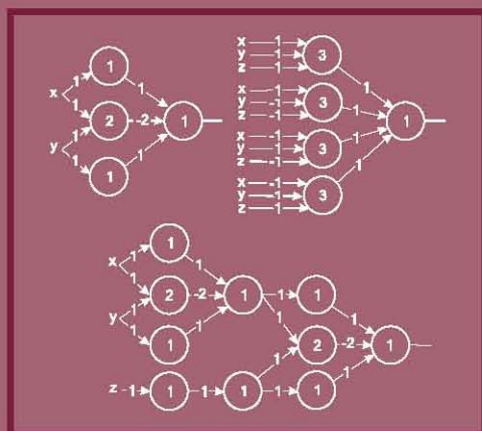
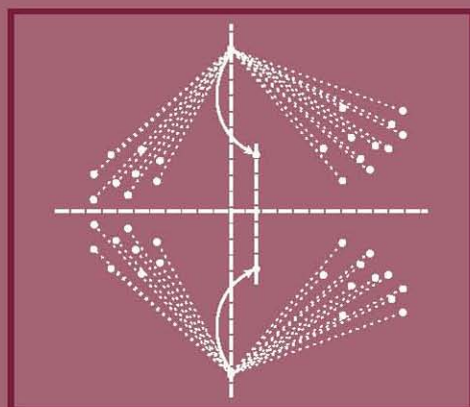


С. И. Николенко
А. Л. Тулупьев



САМООБУЧАЮЩИЕСЯ СИСТЕМЫ



Российская академия наук
Санкт-Петербургский институт информатики и автоматизации РАН

С. И. Николенко, А. Л. Тулупьев

САМООБУЧАЮЩИЕСЯ СИСТЕМЫ

Москва
Издательство МЦНМО
2009

УДК 004.85
ББК 22.18
Н63

Рецензенты:

В. В. Невзоров, д-р физ.-мат. наук, проф. кафедры теории вероятностей и математической статистики математико-механического факультета СПбГУ,

Е. О. Степанов, д-р физ.-мат. наук, зав. кафедрой математического моделирования естественнонаучного факультета СПбГУ ИТМО

Николенко С. И., Тулупьев А. Л.
Н63 Самообучающиеся системы. — М.: МЦНМО, 2009. — 288 с.: 24 илл.

ISBN 978-5-94057-506-1

Книга посвящена одной из самых практически применимых, активных и быстроразвивающихся областей современной информатики, объединяющей множество методов из различных областей математики и не только математики — машинному обучению. В книге обсуждаются основы многих базовых аппаратов машинного обучения: деревья принятия решений, искусственные нейронные сети, генетические алгоритмы, байесовские классификаторы, алгоритмы кластеризации и обучение с подкреплением. Изложение ведётся увлекательным языком, книгу интересно читать, и она доступна даже не очень подготовленному читателю. Однако при этом сохраняется математическая строгость, а наиболее сложные части изложения заинтересуют и профессионалов. Книга снабжена обширной аннотированной библиографией.

Читать книгу смогут даже старшеклассники, хотя она будет представлять несомненный профессиональный интерес и для студентов всех курсов, изучающих математику и информатику, а также для специалистов и аспирантов, ведущих исследования в соответствующих областях. В этом отношении значительная часть материала монографии сможет сыграть роль углублённого учебного пособия.

ББК 22.18

ISBN 978-5-94057-506-1

© МЦНМО, 2009.
© Николенко С. И.,
Тулупьев А. Л., 2009.

Оглавление

Благодарности	7
Введение	11
Об искусственном интеллекте	12
Машинное обучение и интеллектуальные агенты	17
На чём стоит искусственный интеллект	19
Глава 1. Деревья принятия решений	25
§ 1.1. Введение	26
§ 1.2. Структура дерева принятия решений	27
§ 1.3. Энтропия и прирост информации	30
§ 1.4. Деревья принятия решений и булевские функции	37
§ 1.5. Алгоритм ID3	39
§ 1.6. Реализация ID3 на языках Python и Ruby	39
§ 1.7. Проблема критерия прироста информации	46
§ 1.8. Оверфиттинг	48
§ 1.9. Заключение	50
Глава 2. Обучение концептам	53
§ 2.1. Введение	54
§ 2.2. Частичные порядки	55
§ 2.3. Лирическое отступление: функция Мёбиуса	60
§ 2.4. Алгоритм Find-S	68
§ 2.5. Реализация алгоритма Find-S	71
§ 2.6. Алгоритм исключения кандидатов	75
§ 2.7. Заключение	79
Глава 3. Нейронные сети	81
§ 3.1. Введение	83
§ 3.2. Перцептрон	85
§ 3.3. Обучение перцептрона	89

§ 3.4.	Обучение перцептрона на практике	94
§ 3.5.	Метод градиентного спуска	99
§ 3.6.	Нелинейные перцептроны. Сигмоид.	104
§ 3.7.	Алгоритм обратного распространения ошибки	106
§ 3.8.	Реализация нейронной сети на Java	109
§ 3.9.	Заключение	117
Глава 4.	Генетические алгоритмы	119
§ 4.1.	Введение	121
§ 4.2.	Схема генетического алгоритма в деталях	122
§ 4.3.	Генетические операции	124
§ 4.4.	Представление данных	127
§ 4.5.	Отбор	132
§ 4.6.	Дарвин, Ламарк и Болдуин	134
§ 4.7.	Генетические алгоритмы на деревьях	138
§ 4.8.	Генетическое программирование	140
§ 4.9.	Генетическое программирование на практике	145
§ 4.10.	Язык программирования LISP	155
§ 4.11.	Заключение	158
Глава 5.	Байесовское обучение и классификаторы	161
§ 5.1.	Введение	162
§ 5.2.	Теорема Байеса	163
§ 5.3.	Априорные и апостериорные вероятности	168
§ 5.4.	Теорема Байеса, данные и гипотезы	170
§ 5.5.	МАР и задачи классификации	173
§ 5.6.	Оптимальный и гиббсовский классификаторы	175
§ 5.7.	Наивный байесовский классификатор	178
§ 5.8.	Атрибуция текстов	180
§ 5.9.	Байесовское обучение и нейронные сети	189
§ 5.10.	Принцип наименьшей длины описания	191
§ 5.11.	Заключение	193
Глава 6.	Алгоритмы кластеризации	195
§ 6.1.	Введение	196
§ 6.2.	Постановка задачи и виды кластеризации	199
§ 6.3.	Иерархическая кластеризация	202
§ 6.4.	Кластеризация методами теории графов	203
§ 6.5.	Алгоритм ЕМ	208

§ 6.6.	Кластеризация при помощи ЕМ	215
§ 6.7.	Алгоритм k-средних	223
§ 6.8.	Нечёткие алгоритмы кластеризации	230
§ 6.9.	Заключение	232
Глава 7.	Обучение с подкреплением	233
§ 7.1.	Введение	234
§ 7.2.	Как оценивать поведение агента?	236
§ 7.3.	Многорукие бандиты	240
§ 7.4.	Доказуемо оптимальные алгоритмы	241
§ 7.5.	Другие стратегии	246
§ 7.6.	Поиск стратегий в известной модели	248
§ 7.7.	Поиск оптимальных стратегий без модели	250
§ 7.8.	Поиск моделей и оптимальных стратегий по ним	256
§ 7.9.	Игрушечный пример Q-обучения и Дуна	260
§ 7.10.	Заключение	268
Литература		271

Благодарности

Благодарности от С. И. Николенко

Идея этой книги выросла из конспектов лекций вводного курса по машинному обучению. Я читал этот курс сначала в Академии Современного Программирования¹, а затем в Санкт-Петербургском государственном университете информационных технологий, механики и оптики (ИТМО), в котором и сейчас являюсь ассистентом кафедры компьютерных технологий факультета информационных технологий и программирования. Затем этот курс (уже в несколько расширенном виде) читался и в рамках «Computer Science Club»², и в Академическом физико-технологическом университете РАН (АФТУ РАН). Разработка и преподавание курса происходили при непосредственном участии и финансовой поддержке Антона Лиходедова, которому я благодарен в первую очередь.

Разумеется, всё это никогда бы не состоялось, если бы не мой научный руководитель в аспирантуре ПОМИ РАН Эдуард Алексеевич Гирш. Я благодарен ему и руководителю моего дипломного проекта Николаю Александровичу Вавилову — надеюсь, не будет нескромным сказать, что я считаю их своими учителями. Спасибо Александру Шеню, который прочёл рукопись и высказал много ценных замечаний; надеюсь, когда-нибудь у меня получится написать книгу, которую он назовёт хорошей.

В работе над книгой принимали деятельное участие студенты ИТМО; им принадлежит, например, подавляющее большинство примеров программного кода, рассыпанных по этой

¹Бывшая «Академия Борланд», <http://www.borland-academy.ru/>.

²При Санкт-Петербургском отделении Математического Института им. В. А. Стеклова (ПОМИ РАН), <http://logic.pdmi.ras.ru/~infclub/>.

книге. Спасибо Ольге Большаковой, Сергею Вишнякову, Андрею Вокину, Евгению Кирпичёву, Илье Колыхматову, Дмитрию Кочелаеву, Ольге Комалёвой, Тимуру Магомедову, Илье Пименову, Роману Сатюкову, Фёдору Царёву, Антону Яковлеву... уверен, что многих забыл, и искренне прошу прощения. А за возможность работать со всеми этими замечательными людьми — спасибо декану факультета информационных технологий и программирования университета ИТМО Владимиру Глебовичу Парфёнову.

И, конечно, спасибо Вам, читатель, за то, что не поленились заглянуть даже в раздел «Благодарности».

Благодарности от А. Л. Тулупьева

Мой вклад в книгу был бы невозможен, если бы в свои студенческие годы мне не выпала удача слушать лекции и вообще учиться у замечательных преподавателей математико-механического факультета Санкт-Петербургского государственного университета: Николая Кирилловича Косовского, Виктора Григорьевича Кузьменко, Юрия Владимировича Матиясевича, Валерия Борисовича Невзорова, Анатолия Олесьевича Слисенко, Владимира Олеговича Сафонова. Всем им я бесконечно признателен.

Понимание того, что знания передаются с помощью паттернов — а значит, математические модели знаний с неопределённостью должны быть соответствующим образом структурированы, — пришло, когда мне удалось продолжить обучение на факультете социологии университета: с благодарностью вспоминаю, какие интереснейшие курсы по культуральной антропологии и этносоциологии вели Юрий Владимирович Емельянов и Николай Генрихович Скворцов.

Владимир Иванович Городецкий позволил мне присоединиться к его исследованиям в области искусственного интеллекта, познакомил с парадигмой алгебраических байесовских сетей и сформировал на много лет вперёд рациональную программу теоретических исследований в этой области. С исключительной теплотой вспоминаю те годы, когда Владимир Иванович руководил моей диссертационной работой.

Заметная часть материала этой книги сформировалась в результате обзорно-аналитической работы и научных исследований, которые проводились совместно с Анной Камоевной Абрамян, Андреем Владимировичем Лобацевичем, Леонидом Леонидовичем Налчаджи, Дмитрием Александровичем Никитиным, Александром Владимировичем Сироткиным. С радостью пользуюсь возможностью поблагодарить их и за сотрудничество, и за поддержку, и за терпение.

Антон Алексеевич Афанасьев, Максим Владимирович Ментюков, Сергей Сергеевич Синчук, Алексей Михайлович Левин, Дмитрий Михайлович Столяров, будучи студентами математикомеханического факультета, принимали активнейшее участие в лекционных и практических занятиях, на которых обсуждались логико-вероятностные модели знаний с неопределённостью; их вклад существенно повлиял и на содержание соответствующих разделов книги и на форму его изложения.

Введение

Читатель, добравшийся по крайней мере до предисловия и недовольный им, заплатил за книгу деньги и хотел бы знать, какое ему будет возмещение. В этом случае моё последнее прибежище — напомнить ему, что он знает различные способы извлечь пользу из книги без того, чтобы и впрямь её прочитать. Она может, подобно многим другим, прикрыть зазор в его библиотеке, где, аккуратно переплётённая, она наверняка будет хорошо смотреться. Или же он может положить её на туалетный или чайный столик своей учёной подруги. Или же, наконец, он может её отрецензировать — это, без сомнения, лучший способ её употребить, и я особо его рекомендую.

*Мир как воля и представление
Артур Шопенгауэр*

В начале каждой главы её основные события будет предвосхищать Винни. Да-да, тот самый Винни-Пух, который машет вам на рисунке. Истории про Пуха искушённый читатель, конечно, может презрительно пропустить, но мы надеемся, что они будут интересными и достаточно тесно связанными с сутью происходящего. А сейчас Пух просто хочет поздороваться, ведь он действительно очень добрый медвежонок и искренне рад вас видеть.



Об искусственном интеллекте

Идея искусственного интеллекта занимала людей довольно давно. Гефест создавал из металла помощников для своей кузницы, а одного гигантского человекоподобного робота — Талоса — даже подарил Миносу для охраны Крита (по другой версии, Зевс подарил его Европе). Пигмалион создал искусственный интеллект из слоновой кости — правда, без Афродиты у него ничего бы не получилось.

В средние века подобные достижения в основном приписывались магии — например, Вильгельм Парижский писал, что Альберт Великий¹ изготовил голову, говорящую человеческим голосом. Впрочем, были и реальные попытки создать видимость искусственного интеллекта — знаменитый шахматный автомат «Турок» управлялся сидящим внутри карликом-шахматистом, но свою роль в популяризации идеи искусственного интеллекта сыграл. В иудейской традиции искусственный интеллект отражён в виде *големов* — сделанных из глины искусственных слуг, обладающих интеллектом; сделать себе голема могли только самые святые и мудрые. Впрочем, големы умели только понимать речь, но не издавать её: по мнению иудеев, если бы они умели говорить, это бы означало, что у них есть душа. Литературную часть исторического экскурса стоит, пожалуй, закончить доктором Виктором Франкенштейном — дальше идея искусственного интеллекта появляется слишком часто, чтобы подробно её здесь анализировать.

¹ Альберт Великий (Albertus Magnus, ок. 1200–1280) — монах-доминиканец, один из лучших учёных своего времени. Он первым попытался увязать Аристотеля и христианские каноны, развил доктрину свободной воли (на которую много и с удовольствием ссылался, например, Данте), оставил след в логике, ботанике, географии, астрономии (и астрологии, конечно же), минералогии, химии (в частности, ему приписывают открытие мышьяка), зоологии, физиологии, френологии, теории музыки... Разумеется, занятия наукой привели к тому, что Альберта Великого и при жизни, и в особенности после неё считали магом и волшебником. Несмотря на занятия алхимией — Альберт Великий не подтверждает, что нашёл философский камень, но пишет, что считали наблюдал изготовление золота трансмутацией — он был беатифицирован в 1622 году, а в 1931 канонизирован. Фома Аквинский был учеником Альберта Великого.

Первое полноценное философское обоснование мечты об искусственном интеллекте получили в рационалистических работах Рене Декарта¹. Декарт считал, что в сознании мыслящих людей от рождения заложены теоретические идеи, из которых всевозможные знания получаются сугубо дедуктивным путём, от общего к частному; в этом и есть суть рационализма, в отличие от эмпиризма, который предпочитает идти индуктивно, от частного к общему. Отсюда уже полшага до искусственного интеллекта: дедукцию гораздо легче автоматизировать, чем индукцию (хотя и на этом пути есть свои проблемы), а дальше нужно будет только выделить и записать эти «основные теоретические идеи». Декарту вообще было свойственно пытаться механизировать человека и человеческие рассуждения: в известной цитате он сравнивает больного человека со сломанными часами, а одной из главных задач декартовой аналитической геометрии была попытка автоматизировать, упростить красивые индуктивные геометрические построения, которыми до той поры приходилось доказывать теоремы.

Поворотным моментом в истории искусственного интеллекта стало, разумеется, изобретение первых компьютеров. Собственно, история искусственного интеллекта как раздела информатики начинается с начала 1950-х гг. — в 1950 году Алан Тьюринг² предложил свой знаменитый тест, который считается

¹Рене Декарт (René Descartes, 1596–1650) — французский учёный, философ, математик, писатель и многое-многое другое. Декарт родился во Франции, выучился на юриста, а затем поступил на военную службу в революционной Голландии, где впервые заинтересовался математикой. Именно в Голландии прошли наиболее плодотворные годы Декарта: с 1628 по 1649 он написал подавляющее большинство своих работ. А работы эти фактически заложили фундамент всей современной науки: в «Рассуждении о методе» Декарт, основываясь на скептическом методе, создаёт рационалистическую систему познания, основанную на дедуктивных рассуждениях, а также разрабатывает дуалистическую теорию, в которой «душа», мышление отделяются от тела и рассматриваются как внефизические процессы. В математике Декарт известен как автор аналитической геометрии; кроме того, именно он впервые применил анализ бесконечно малых к задаче о касательной, что позволило Ньютону и Лейбницу разработать основы математического анализа.

²Алан Тьюринг (Alan Turing, 1912–1954) — английский математик, логик и криптограф. Он во многом создал современную теоретическую информатику, формализовав понятие алгоритма, введя в рассмотрение машины Тьюринга и сформулировав общепринятую версию тезиса Чёрча (в

отправной точкой AI¹ как науки [138, 159]. Суть теста Тьюринга в том, что компьютер должен суметь успешно выдать себя за человека в (письменном) диалоге между судьёй-человеком и компьютером: судья одновременно ведёт два диалога, один с настоящим человеком, а другой с компьютером, и после этого должен суметь ответить, кто из них кто. Сам Тьюринг, впрочем, тут же указывал, что наоборот у нас, людей, уже никак не получится: человека от компьютера легко отличит любая достаточно сложная арифметическая задачка. Отметим здесь же, что в статье самого Тьюринга есть некоторые детали, которые обычно ускользают из современных формулировок. Например, Тьюринг предлагал сравнивать не компьютер, имитирующий человека, и настоящего человека, а компьютер, выдающий себя за женщину, и мужчину, выдающего себя за женщину; иначе говоря, сравнивать *имитирующий кого-то другого* компьютер с *имитирующим кого-то другого* человеком (подробнее о тесте Тьюринга см. [138]).

Пятидесятые–шестидесятые годы XX века были временем безудержного оптимизма. Казалось, что ещё чуть-чуть, ещё немного — и гигантские холлы, заставленные громоздкими ламповыми устройствами, напечатают на перфокартах: «Кто я?». Научная фантастика того времени (которое заслуженно считается «золотым веком» научной фантастики вообще) уже не просто мечтает об искусственном интеллекте, а считает его свершившимся фактом, делом недалёкого будущего. Мысль движется дальше: фантасты задаются вопросами о будущем сосуществовании человека и разумной машины. Всем известны три закона

англоязычной литературе называемого обычно Church–Turing thesis). Именно Тьюринг переформулировал теорему Гёделя о неполноте, применив её к гильбертовской «Entscheidungsproblem», то есть к задаче о том, можно ли алгоритмически искать доказательства математических утверждений. Кроме того, он был одним из отцов-основателей современного искусственного интеллекта; ему принадлежит формулировка «теста Тьюринга», который должна пройти машина, претендующая на право считаться «искусственным интеллектом»: суметь поддержать разговор (хотя бы в текстовом формате) с человеком на естественном языке так, чтобы человек не понял, машина или другой человек с ним разговаривает.

¹AI — общепринятое сокращение от Artificial Intelligence, искусственный интеллект.

робототехники, сформулированные Айзеком Азимовым¹; они, если вдуматься, предельно неформальны, понять их и принять к исполнению — тест гораздо сильнее тьюринговского.

Литературный и общественный оптимизм отразился и в науке. Пятидесятые годы — время расцвета математической логики; несмотря на уже доказанную и осознанную теорему Гёделя о неполноте, для многих ограниченных логик уже были известны разрешающие процедуры. Да, они бывали экспоненциальными, дважды экспоненциальными и даже более того, но общий вектор развития виделся ясным — автоматический логический вывод казался делом ближайшего будущего, а на этой основе можно было строить и ещё более амбициозные планы.

В 1956 году Джон Маккарти², Марвин Мински³, Натаниэль Рочестер⁴ и Клод Шеннон организовали знаменитую летнюю школу в Дартмуте, где мозговым штурмом попытались

¹ Айзек Азимов (Isaac Asimov, 1920–1992) — американский фантаст и популяризатор науки. Родился Азимов в деревне Петровичи Могилёвской губернии, ныне Смоленской области (звали его тогда Исаак Озимов, от слова «озимые», которыми торговал его прадед). Хотя Азимов главным образом известен как фантаст — его достижения в этой сфере мы здесь не будем и пытаться перечислять — он также написал множество научно-популярных книг, от «Краткой истории химии» до «Путеводителя Азимова по Шекспиру». Одной из главных тем творчества Азимова стало взаимодействие человека с мыслящими машинами, для которого он сформулировал свои знаменитые три закона.

² Джон Маккарти (John McCarthy, p. 1927) — американский информатик и когнитивист. Маккарти придумал термин «artificial intelligence», а также создал язык программирования Lisp, ставший для раннего AI основным [104]. Кроме того, Маккарти много работал в математической логике, в частности, над немонотонными логиками [105].

³ Марвин Мински (Marvin Minsky, p. 1927) — американский исследователь в области искусственного интеллекта, один из основателей лаборатории AI в Массачусетском технологическом институте, автор классической книги о нейронных сетях «Перцептроны» [111]. Примечательно, что Мински консультировал Артура Кларка при написании и постановке его «Космической Одиссеи 2001» и даже весьма лестно там упоминается — см. эпиграф к главе 3; цитата, кстати, весьма характерная для тех оптимистических времён.

⁴ Натаниэль Рочестер (Nathaniel Rochester, 1919–2001) — американский инженер и математик. Рочестер стоял у истоков не только искусственного интеллекта, но и других областей: в компании IBM в начале пятидесятых он создал первый компьютер компании IBM 701, а в шестидесятые годы занимался криогеникой.

заложить основы будущей теории искусственного интеллекта. В заявке на спонсирование этой летней школы Маккарти писал [106]:

Мы предлагаем исследование искусственного интеллекта сроком на два месяца с участием десяти человек летом 1956 года в Дартмутском колледже, Гановер, Нью-Гемпшир. Исследование основано на предположении, что всякий аспект обучения или любое другое свойство интеллекта может в принципе быть столь точно описано, что машина сможет его имитировать. Мы попытаемся понять, как обучить машины использовать естественные языки, формировать абстракции и концепции, решать задачи, сейчас подвластные только людям, и улучшать самих себя. Мы считаем, что существенное продвижение в одной или более из этих проблем вполне возможно, если специально подобранная группа учёных будет работать над этим в течение лета.

Вот такой незамысловатый проект — по сути ерунда, работы немного, десятку человек на одно лето... В реальности же оказалось, что искусственный интеллект — задача почти безнадежная. Несмотря на многолетние усилия лучших умов мира, создать пусть даже не лучший из умов пока не получается. Какой там тест Тьюринга — даже простое чтение безо всякого понимания оказывается для компьютера непосильной задачей, о чём свидетельствуют многочисленные CAPTCHA-тесты на различных сайтах¹.

Проблемы с «человекоподобными» задачами появляются даже в настольных играх, где, на первый взгляд, задача достаточно хорошо формализована. Компьютер может хорошо играть в шахматы, даже обыгрывать за доской ведущих гроссмейстеров

¹CAPTCHA (Completely Automated Public Turing Test to tell Computers and Humans Apart) — это тест, в котором от человека требуется распознать написанный текст или какое-нибудь другое свойство предъявленного ему изображения. Такие тесты часто используются для «защиты от ботов», — доказательства того, что тест проходит живой человек, а не программа (бот) [1, 2].

(на сегодня рейтинг Эло лучших компьютеров составляет около 3100–3200, а лучших живых шахматистов — 2700–2800), потому что в шахматах не слишком большая разветвлённость дерева поиска даёт большой простор для применения вычислительных мощностей. Но, например, в игре го, где нужно значительно тоньше понимать позицию на всей доске и где умение быстро считать варианты не может радикально усилить игру, компьютеры пока не выдерживают никакой критики. Лучшие компьютерные программы играют в го на уровне не слишком увлечённого любителя, а профессионалам проигрывают с гигантской форой, доходящей до 20–25 камней (кому может чемпион мира по шахматам дать ферзя и двух ладей вперёд?) [23, 115]. И это — строго формализованная игра на конечной, весьма ограниченной доске¹. О какой-либо деятельности, где нужно понимать естественный язык, конечно, и говорить не приходится.

Однако поле искусственного интеллекта весьма широко, и было бы в корне неправильно оценивать успехи отрасли сугубо бинарно, по факту достижения основной цели. На пути к ней были разработаны многие подходы, которые получили широкое распространение и оказались очень полезны на практике. Большая часть этих подходов объединена под названием *машинного обучения* (*machine learning*). Именно ему посвящена эта книга.

Машинное обучение и интеллектуальные агенты

Что такое машинное обучение? Общее (хотя и не слишком формальное) определение таково: алгоритм является алгоритмом машинного обучения, если он улучшает своё поведение по мере накопления опыта. Что это значит? Это значит, что алгоритм *обучает* параметры модели либо на заранее подготовленных тестовых примерах, либо на собственных ошибках, и со временем решает поставленную задачу всё лучше и лучше. Некоторые алгоритмы машинного обучения способны подмечать ранее неизвестные закономерности в данных, выделять *знания* (тоже

¹Позднейший комментарий: к 2008 году наметился прогресс и в го; программа MoGo, основанная на специальном алгоритме построения дерева поиска, смогла вывести компьютерное го на новый уровень; правда, о мировом господстве пока мечтать рано, но уровень где-то первого дана достигнут [46, 47].

термин и предмет искусственного интеллекта), которых раньше не было — словом, делать то, что долгое время считалось прерогативой исключительно *Homo Sapiens*.

Обучающиеся агенты — частный случай интеллектуальных агентов, которые изменяют своё поведение в зависимости от реакции окружающей среды. Главным отличием обучающихся агентов от всех остальных является то, что они не просто действуют по какой-либо программе, в которую заранее заложена реакция на разные действия среды, а *улучшают* своё поведение *с опытом*. Иными словами, в определении алгоритма машинного обучения всегда присутствует требование повышать эффективность по ходу работы или при увеличении начальной обучающей выборки.

Обучающиеся агенты всегда действуют в условиях либо неполной информации, либо информации, полностью обработать которую вычислительно нереально — иначе можно было бы просто выбирать оптимальное поведение, и обучаться было бы не нужно. А нетривиальные алгоритмы для обучающихся агентов появляются только тогда, когда оптимальное со статистической точки зрения поведение слишком трудно вычислить полным перебором. Приведём здесь же ссылки на главные источники по машинному обучению — книги [5, 19, 63, 92, 101, 107–109, 113].

На рисунке 1.1 изображена схема (очень общая и приближенная) самообучающегося агента. Давайте рассмотрим основные её элементы:

- *решатель* — мозг самообучающегося агента, именно он получает информацию из внешнего мира и выбирает, какие действия производить;
- *критик* связывается с неким стандартом (оценкой, идеалом) эффективности и определяет, насколько хорошо отработал агент;
- *обучающийся элемент* улучшает поведение агента, изменяя его параметры в соответствии с тем, хорошо ли получилось у решателя (о том, что такое хорошо и что такое плохо, этой крохе сообщит критик); он же сообщает генератору задач, какой ещё нужен опыт для дальнейшего обучения;

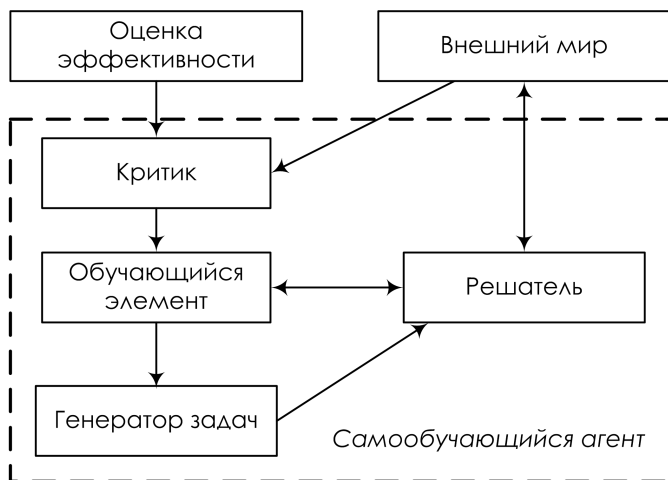


Рис. 1.1. Схема самообучающегося агента

- *генератор задач* предлагает действия, которые привели бы к новому, полезному опыту: во многих областях применения машинного обучения агенту приходится выбирать, какие ещё стратегии исследовать (а может, удовольствоваться тем, что он уже знает).

Не во всех аппаратах машинного обучения присутствуют все эти элементы, и во многих из них несколько элементов схемы объединяются в один. Однако схема помогает понять, как действует большинство алгоритмов машинного обучения.

На чём стоит искусственный интеллект

Искусственный интеллект вообще и машинное обучение в частности — поле в высшей степени междисциплинарное. В нём много раз использовались идеи из разных областей знания, на первый взгляд совершенно не похожих ни друг на друга, ни на информатику. В книге мы неоднократно будем указывать на подобные обстоятельства, а сейчас просто вкратце опишем главное: на чём стоит современный искусственный интеллект, фундамент, без которого его трудно было бы себе представить.

Математика, конечно, в любом случае остаётся царицей и служанкой любой науки. А поскольку искусственный интеллект — раздел информатики, в котором речь идёт об идеализированных объектах (функциях, алгоритмах, оценках сложности, структурах данных), без математики здесь никуда. Однако некоторые разделы математики встречаются в искусственном интеллекте значительно чаще, чем другие.

Теорию вероятностей и статистику можно без преувеличения назвать основой всего машинного обучения вообще и значительной доли других исследований в рамках искусственного интеллекта. Даже если алгоритм на первый взгляд не использует вероятности или случайные процессы, при ближайшем рассмотрении наверняка окажется, что для его анализа придётся привлекать вероятность. Об этом мы будем подробно говорить в главе 5, да и во всех остальных.

Математическая логика пронизывает всю информатику. Бинарная логика, созданная Джорджем Булем¹, стала основным инструментом информатики. Да и вообще информатика была во многом мотивирована логикой — оптимизм логицистов начала XX века, веривших, что всю математику можно автоматически вывести из умело подобранных аксиом, был одной из отправных точек развития автоматического вывода. Апофеозом логицизма стал многотомный труд Альберта Уайтхеда и Бертра-на Рассела «Principia Mathematica» («Принципы математики»), в котором из аксиом выводились арифметические утверждения. Книга эта очень похожа на автоматически порождённый логический вывод, и к странице 379 первого издания авторы наконец-то заключают: «Когда мы определим арифметическое сложение, из этого предложения будет вытекать, что $1 + 1 = 2$ ».

¹Джордж Буль (George Boole, 1815–1864) — английский математик и философ, один из основателей информатики. Трудно рассказать о Буле какой-нибудь забавный случай, потому что жизнь его была на редкость стабильной и безынтересной — он всю жизнь проработал преподавателем математики, сначала в школе, потом в колледже. Трудом его жизни стал двухтомник «Исследование законов мышления, на которых основаны математические теории логики и вероятностей» [20], где Буль исследовал связь между алгеброй, логикой и вероятностью и развил законы алгебраической записи логических законов и бинарной логики. Идеи Буля в XIX веке никого не заинтересовали — потребовался Клод Шеннон, чтобы понять, что булева логика может стать основой информатики.

Правда, оптимизма логицистам изрядно поубавили теоремы Гёделя¹ о неполноте, но задача автоматизации логического вывода с повестки дня не сошла. Она и по сей день остаётся одним из краеугольных камней не только теории доказательств, но и искусственного интеллекта: многие аппараты должны проводить логический вывод (или его аналоги). Кроме того, искусственный интеллект часто работает с различными мерами неопределённости [57, 176, 177]. В этом ему помогают вероятностные логики [176], основы которых заложил ещё Буль [20], и нечёткие логики [40, 117].

Теория игр изучает взаимодействие агентов, каждый из которых стремится выбрать стратегию поведения, оптимальную с точки зрения той или иной целевой функции. Разумеется, такого рода задачи тесно связаны с задачами машинного обучения. Теория игр изучает те случаи, в которых машинное обучение уже не обязательно: она предоставляет готовые оптимальные рецепты в той или иной ситуации. Однако зачастую, прежде чем использовать эти рецепты, нужно установить некоторые параметры ситуации, в которой находится самообучающийся агент, или задать принципы взаимодействия нескольких агентов. Поэтому теория игр активно используется в алгоритмах машинного обучения. Примерно ту же роль играет и *теория управления*, которая тоже решает задачи максимизации заданной функции.

¹Курт Гёдель (Kurt Gödel, 1906–1978) — австрийский математик. В 1924 г. поступил в Венский университет, в 1930 г. защитил диссертацию по математике; в своей диссертации Гёдель доказал полноту исчисления предикатов первой ступени, что дало надежду на доказательство непротиворечивости и полноты всей математики. Однако уже в 1931 г. Гёдель надежду отнял, доказав свою знаменитую теорему о неполноте. Согласно этой теореме, любая процедура доказательства истинных утверждений арифметики обречена на неполноту. Следовательно, внутренняя непротиворечивость любой математической теории не может быть доказана иначе, как с помощью обращения к другой теории, использующей более сильные аксиомы, а значит, менее надёжной. Методы, использованные Гёделем при доказательстве теоремы о неполноте, сыграли в дальнейшем важную роль в информатике. Правда, позитивных утверждений у Гёделя было всё-таки больше: например, в 1938 г. он доказал, что присоединение аксиомы выбора и континуум-гипотезы к обычным аксиомам теории множеств не приводит к противоречию.

Научное издание

Утверждено к печати Учёным советом
Санкт-Петербургского института информатики и автоматизации РАН

Сергей Игоревич Николенко
Александр Львович Тулупьев

САМООБУЧАЮЩИЕСЯ СИСТЕМЫ

Оригинал-макет подготовлен в издательской системе
L^AT_EX_{2 ϵ}

Художник:	<i>Ефремова Н. С.</i>
Редактор:	<i>Коноваленко Е. А.</i>
Компьютерная вёрстка:	<i>Николенко С. И.</i>

Подписано в печать 20.04.09 г. Формат 60×90/16. Бумага офсетная.
Печать офсетная. 18 печ. л. Тираж 1000 экз. Зак. № .

Издательство Московского центра непрерывного математического
образования. 119002, Москва, Большой Власьевский пер., 11.
Тел. (499) 241-74-83.

Отпечатано с готовых диапозитивов в ППП «Типография "Наука"».
121099, Москва, Шубинский пер., 6.

Книги издательства МЦНМО можно приобрести в магазине
«Математическая книга», Москва, Большой Власьевский пер., 11.
Тел. (499) 241-72-85. E-mail: biblio@mccme.ru