А. С. Потапов

РАСПОЗНАВАНИЕ ОБРАЗОВ и МАШИННОЕ ВОСПРИЯТИЕ

Электронный аналог печатного издания: Потапов А. С. Распознавание образов и машинное восприятие: Общий подход на основе принципа минимальной длины описания. — СПб. : Политехника, 2007. — 548 с. : ил.

УДК 004.855
ББК 32.973.26
П64

Рецензенты: член-корреспондент РАН, доктор технических наук, профессор М. М. Мирошников и кафедра компьютерной фотоники Санкт-Петербургского государственного университета информационных технологий, механики и оптики

Потапов А. С.

П64 Распознавание образов и машинное восприятие: Общий подход на основе принципа минимальной длины описания. — СПб.: Политехника, 2011. — 548 с.: ил.

ISBN 5-7325-0881-3

В книге подробно рассмотрен принцип минимальной длины описания, являющийся следствием теоретико-информационного подхода к построению моделей и выбору гипотез. Этот принцип становится все более популярным при решении сложных задач автоматического анализа данных, традиционно относившихся к области искусственного интеллекта. Рассмотрены задачи распознавания образов, машинного восприятия и грамматического и логического выводов, для которых использование принципа минимальной длины описания уже позволило получить более эффективные решения. На конкретных примерах показана возможность разработки унифицированного подхода к решению указанных задач.

Книга предназначена для широкого круга читателей: студентов, молодых ученых и специалистов, интересующихся компьютерными науками и, в частности, искусственным интеллектом.

УДК 004.855
ББК 32.973.26

ISBN 5-7325-0881-3

ОГЛАВЛЕНИЕ

Глава 1
ИНДУКТИВНЫЙ ВЫВОД

1.1. ПРОБЛЕМА ВЫБОРА ГИПОТЕЗ В ИНДУКТИВНОМ ВЫВОДЕ

1.1.1. Что такое индуктивный вывод? Неформальное рассмотрение

Все рациональные рассуждения традиционно делятся на дедуктивные и индуктивные [1, с. 141]. Принято считать, что индукция — это умозаключение от частных фактов к некоторому общему гипотетическому утверждению, в то время как дедукция — это способ рассуждения, при котором осуществляется переход от общего знания или фактов к частным следствиям. Однако индуктивному выводу придается и более широкий смысл, если рассмотрение не ограничивается формальной логикой. Наиболее широко индуктивный вывод можно определить как проблему выбора модели из некоторого множества моделей, которая наилучшим образом «объясняет» исходные данные [2, с. 1]. Здесь под частными фактами понимается набор данных, а под общим утверждением — модель, описывающая эти данные (содержащиеся в них закономерности).

Это означает, что индуктивным выводом будет являться и проведение некоторой интерполяционной кривой по заданному набору точек, и составление словесного описания изображения. Более того, многие виды реальных рассуждений, традиционно относимых к дедуктивным, могут также быть причислены и к индуктивным. Так, классический пример дедуктивного рассуждения [1, с. 143]: «Все люди смертны. Сократ — человек, следовательно, Сократ смертен» опирается на две посылки, по крайней мере, первая из которых («Все люди смертны») не является с необходимостью истинной, а является результатом обобщения данных опыта. Тогда и консеквенту (заключению) этого вывода («Сократ смертен») можно присвоить лишь некоторую вероятность, отличную от единицы, а значит, в этом выводе производится выбор более достоверной гипотезы из двух возможных. Именно недостоверность результата и является признаком индуктивного вывода, отличающим его от де-

дуктивного вывода, в котором следствие с необходимостью получается из посылок и истинность посылок переносится на следствие.

Индуктивные рассуждения являются неотъемлемой частью естествознания, а недостоверность индуктивного вывода тесно связана с проблемой обоснования научного знания, которая наиболее отчетливо проявилась в философии Нового времени. В связи с этим изучение индукции изначально проводилось в философии науки. Попытки разработать адекватную логическую теорию индуктивного вывода (или индуктивную логику) проводились со времен Фрэнсиса Бэкона. Однако классическое понимание индукции как простого обобщения эмпирических фактов (результатов наблюдений, физических измерений или экспериментов над объектами внешнего мира) приводит к непреодолимым трудностям. Еще Д. Беркли заметил, что на основе самого по себе индуктивного подхода идеализм, в частности субъективный, неопровержим, поскольку невозможно установить, над чем именно осуществляется исходное наблюдение [3, с. 360]. И действительно, невозможно отличить феномен, даваемый нам в виде каких-то ощущений, от феномена, который совпадает с самими этими ощущениями.

Более детально эта проблема была рассмотрена Дэвидом Юмом, который впервые подверг глубокому исследованию понятие причинности. Так, в феноменологическом эмпиризме Юма как причинно-следственная связь, так и общие понятия являются не более чем психологической привычкой к ассоциативному связыванию идей («копий» с первоначальных впечатлений). Связывание идей оказывается возможным лишь как результат деятельности мышления и не зависит от наличия объективного аналога итога такого связывания. Выявленные слабости существовавшего к тому времени понимания индуктивной логики вообще поставили под сомнение ее право называться логикой. В результате Юмом была в общем виде сформулирована следующая задача [3, с. 363]: дать строгое, точное и объективное обоснование и оправдание индуктивной логики. Эта задача до сих пор не решена.

Поскольку чисто эмпирический подход к индуктивной логике, т. е. ее рассмотрение как простое обобщение результатов наблюдений и экспериментов, потерпел неудачу, то стало ясно, что она должна базироваться на некотором более прочном фундаменте. Возможно, именно поэтому Кант пытался построить метафизическую теорию, чтобы найти для

строго научного познания сущности изучаемых феноменов априорные основания, лежащие вне сферы чувственного опыта [3, с. 364]. Проблемы, родственные кантовскому вопросу о том, как возможны априорные синтетические суждения, сейчас заново встают при разработке систем машинного обучения — области знаний, тесно связанной с индуктивным выводом и являющейся разделом искусственного интеллекта (часто также называемого «практической гносеологией»).

Другой родственной областью исследований является автоматизация научных исследований, в которой исследуются проблемы выдвижения и проверки гипотез (см., например, [4]). Эта область знаний лежит на стыке философии науки и искусственного интеллекта (как научной дисциплины), в котором проблема индуктивного вывода играет существенную роль. Поэтому часто вопрос «Может ли машина мыслить?» конкретизируется как «Может ли машина формулировать и проверять гипотезы?» [5, с. 11].

Именно в искусственном интеллекте проблема индуктивного вывода получила свое дальнейшее развитие, поскольку эта область предоставляет наиболее сложные и интересные задачи, с одной стороны, и требует явного задания правил вывода (многие из которых осуществляются человеком неосознанно) — с другой.

И наконец, еще одно направление исследований, непосредственно связанное с индуктивным выводом, — статистический анализ, в котором также производится построение модели данных. Таким образом, есть несколько проблем, близких по содержанию к индуктивному выводу, — это машинное обучение, автоматическое выдвижение научных гипотез и статистический анализ данных. Все они могут быть охарактеризованы как вычислительный индуктивный вывод, который и будет в центре нашего внимания. Поскольку указанные три направления относятся к разным областям науки и философии, используемая в них терминология несколько различается. В связи с этим основные понятия требуют определенного уточнения.

1.1.2. Основные понятия индуктивного вывода

Рассмотрим некоторые важнейшие понятия индуктивного вывода, которые будут являться центральными для всего дальнейшего изложения. При этом будем учитывать, что

для каждого из этих понятий существует набор эквивалентных терминов, свойственных различным областям знаний.

• Исходная информация, на основе которой осуществляется вывод, может обозначаться такими терминами, как «частные факты», «набор исходных данных» («данные наблюдений»), «выборка измеренных значений случайной величины», «реализация случайного процесса». Здесь будет использоваться в основном термин «данные», а термин «выборка» будет фигурировать преимущественно в контексте рассмотрения статистических методов. Исходный набор данных в дальнейшем будет обозначаться символом D.

• Для обозначения результатов индуктивного вывода могут использоваться следующие понятия: «гипотеза» (реже — «теория»), «общее правило», «модель» и «оцененные параметры». Термины «гипотеза» и «модель» мы будем употреблять наравне, привнося в них в качестве отличия лишь слабый смысловой оттенок следующего содержания. В определенных контекстах под гипотезой будет пониматься атомарный объект, принадлежащий некоторому множеству таких же неделимых объектов, а под моделью — объект, обладающий (возможно, сложной) внутренней структурой. Таким образом, в то время как гипотезы выбираются, модели могут конструироваться. Некоторая гипотеза будет обозначаться с помощью символа h, если сущность гипотезы не уточняется (например, если не говорится, что гипотезой является программа для машины Тьюринга или класс образов).

• Все возможные результаты вывода объединяются в пространство (или множество) гипотез либо образуют класс моделей. Наряду с этими терминами может также использоваться и такое понятие, как «язык представления» (или просто «представление»). Это понятие особенно удобно использовать, когда модель, описывающая исходные данные, задается в виде цепочки символов. Изредка может использоваться также и такой термин, как «метамодель», который указывает, что само пространство гипотез (или язык представления) может варьироваться, являясь результатом индуктивного вывода следующего уровня, имеющего дело с целой предметной областью. Пространство гипотез будет обозначаться с помощью символа H.

• Одним из наиболее важных элементов индуктивного вывода является критерий, с помощью которого производится сравнение альтернативных гипотез. Он также называ-

ется критерием рациональности выводов и может уточняться либо как точность предсказания, даваемая моделью, либо как близость данной модели к «истинной» модели. На возможных вариантах задания этого критерия мы чуть подробнее остановимся ниже. Качество некоторой гипотезы h при условии, что есть исходные данные D, будет обозначаться как $r(h \mid D)$.

• Для наименования самого процесса вывода могут использоваться различные термины, конкретизирующие привлекаемый метод, например «статистический вывод» или, еще более узко, «байесовский вывод». Более общие названия этого процесса: индуктивный вывод, оценивание параметров, выбор или поиск модели. Выбор лучшей гипотезы можно описать следующим образом:

$$h^* = \arg\max_{h \in \mathrm{H}} r(h \mid D). \qquad (1.1)$$

Здесь не указывается, к каким именно предметным областям относятся те или иные термины. Такую информацию можно найти, например, в работе [2, табл. 1.1–1.3].

Помимо терминологических расхождений в различных подобластях индуктивного вывода дополнительная путаница может возникать из-за существования определенных отличий индуктивного вывода от близких проблем анализа данных, в которых ставятся другие цели, такие как предсказание или принятие решений (о вопросе разделения индуктивного вывода и теории принятия решений см., например, [1, гл. 13–14]). Для нашего изложения эти различия зачастую будут несущественными, но при необходимости мы будем на них указывать.

1.1.3. Критерии сравнения гипотез

Сформулировав задачу индуктивного вывода как выбор из некоторого множества модели, наилучшим образом объясняющей исходные данные, приходим к первичной проблеме, заключающейся в установлении приемлемого критерия для выбора лучшей модели. Нахождение такого критерия — это центральный вопрос, общий для таких областей, как статистический анализ, машинное обучение и философия науки [2, с. 3]. Отметим, что здесь идет речь именно об универсальном критерии, который можно было бы

использовать при решении любой задачи, представляемой в виде индуктивного вывода. Частные же критерии, которые придумываются человеком для решения конкретных задач, обычно оказываются неприменимы в новых задачах, поэтому для решения последних требуется творческое участие человека.

Было бы естественно предположить, что лучшая модель — это та, которая наиболее близка к истинной модели. Но тогда нужно было бы не только иметь возможность задать метрику в пространстве моделей, но и заранее знать истинную модель, а это доступно лишь в исключительных случаях. В некоторых задачах статистического анализа вводятся частные критерии близости данной модели к истинной, такие, как, например, среднеквадратичное отклонение. Однако подобные критерии, хотя и могут казаться интуитивно очевидными, перестают давать адекватный результат как только нарушаются заложенные в них (явные или неявные) априорные предположения. Такие примеры мы еще подробно разберем.

Принципиально другой подход к обсуждаемой проблеме заключается в выборе той модели, которая дает наибольшую точность предсказания. Классическим приемом для получения объективной оценки точности предсказания является разделение выборки на обучающую и тестовую части. Существуют также различные методы перекрестной проверки. Однако, во-первых, использование тестовой выборки приводит к уменьшению объема данных, по которым строится модель, а значит, понижается и точность модели. Во-вторых, не во всех задачах индуктивного вывода можно численно выразить точность предсказания. Поэтому ее желательно оценивать косвенно, привлекая некий другой критерий.

В философии науки используются такие критерии, как простота гипотезы (часто, особенно в зарубежной литературе, этот критерий связывается с принципом бритвы Оккама) и ее фальсифицируемость (отождествляемая с содержательной емкостью) [1, с. 231]. Принцип фальсифицируемости, введенный К. Р. Поппером, гласит, что выбирать нужно ту гипотезу, которая раньше других опровергалась бы новыми данными, полученными в результате наблюдений или эксперимента, если была бы ложной. Возможно, что понятия простоты и фальсифицируемости по смыслу достаточно близки [1, с. 233]. К сожалению, эти критерии ос-

таются бесполезными для вычислительного индуктивного вывода до тех пор, пока не являются вполне формализованными.

Именно понятие простоты используется в байесовских методах для определения априорных вероятностей моделей [5, с. 715]. В этих методах (как и в ряде других статистических методов) лучшей считается наиболее вероятная модель. Вообще, понятие вероятности неразрывно связано с индуктивным выводом: «...не только при анализе статистических выводов, но и при обсуждении, на первый взгляд, чисто качественных проблем индукции исчисление вероятностей играет центральную роль. Более того, хотя статистические выводы можно считать всего лишь частными и нетипичными образцами индуктивных выводов, нельзя сколько-нибудь обоснованно отказать им в принадлежности к области индуктивной логики» [1, с. 6]. А поскольку за байесовским выводом закрепилась репутация оптимального вывода, то следующая глава будет посвящена его рассмотрению.

Проблемой, сопутствующей установлению критерия рациональности гипотез, является выбор пространства гипотез, размер которого может заметно варьироваться в зависимости от задачи. Так, в статистическом выводе могут рассматриваться однопараметрические классы моделей, а могут — и гораздо большей размерности. Но в целом статистический анализ характеризуется наиболее ограниченными пространствами гипотез. В машинном обучении существуют проблемы, тесно примыкающие к статистическому выводу и также вовлекающие пространства гипотез «обозримого» размера. Наименее ограниченные пространства гипотез рассматриваются, пожалуй, в индуктивном выводе, изучаемом философией (что хорошо видно по возникающим здесь парадоксам, на которых мы еще остановимся позднее), и при разработке универсальных систем машинного обучения.

Ограничение, накладываемое на пространство гипотез, можно трактовать как априорно принятое решение об отказе проводить сравнение качества всех гипотез, не вошедших в выбранное пространство. Таким образом, задание пространства гипотез и определение критерия их сравнения — это разные стороны одной и той же проблемы, что более явно будет показано позднее. Поэтому неудивительно, что наибольшие сложности в установлении приемлемого критерия сравнения моделей возникают именно в фило-

софии науки и при разработке «сильного» искусственного интеллекта. Хотя эти теоретические сложности и находят свое отражение в конкретных практических проблемах, но последние менее ярко выражены. Принцип бритвы Оккама, привлекаемый в первой из этих областей, нашел во второй области свое формальное численное воплощение, которое позволяет разрешить эти сложности, по крайней мере частично. Оно и будет составлять основной предмет данной книги.

1.1.4. Бритва Оккама и принцип минимальной длины описания

Простота гипотезы — это один из наиболее часто применяемых критериев в индуктивном выводе (см., например [1, гл. 12]). Однако сама по себе простота гипотезы не может являться критерием при выборе модели, поскольку самая простая гипотеза — это просто отсутствие какой-либо регулярной модели, выявляющей внутренние закономерности в данных. Так, на любой наблюдаемый факт мы можем сказать: «Такова божья воля». Другими словами, простейшая гипотеза гласит, что данные абсолютно случайны, что, естественно, допускает и произвольную экстраполяцию, а значит, никак не может помочь в прогнозировании.

С другой стороны, точность, с которой модель описывает данные, тоже не является подходящим самостоятельным критерием. И действительно, существуют так называемые *гипотезы ad hoc*, которые просто повторяют имеющиеся данные, производят описание данных без их объяснения. Подобные гипотезы ad hoc также не могут помочь в прогнозировании. При этом они абсолютно точно описывают данные, но обладают значительной по сравнению с простейшей гипотезой сложностью.

В связи с этим принято считать, что лучшая гипотеза, дающая наибольшую точность предсказания, — это компромисс между простотой гипотезы и тем, насколько хорошо она удовлетворяет данным наблюдений [5, с. 715]. В философии науки такое положение часто связывается с принципом бритвы Оккама [2, с. 10]. Этот принцип гласит: «То, что можно объяснить посредством меньшего, не следует выражать посредством большего» (Frustra fit per plura quod potest fieri per pauciora), или «Без необходимости не следу-

ет утверждать многое» (Pluralitas non est ponenda sine necessitate). Чаще приводится другая формулировка: «Сущностей не следует умножать без необходимости» (Entia non sunt multiplicanda sine necessitate), но она, по-видимому, в произведениях Уильяма Оккама не встречается.

Как уже было замечено, в байесовских методах простота гипотезы используется для задания априорных вероятностей гипотез, и такие методы также называют формализацией бритвы Оккама [5, с. 715; 6, п. 1.1]. Однако более интересным вариантом формализации понятия простоты, с нашей точки зрения, оказываются подходы, основанные на теории информации. Здесь сложность (как противоположность простоты) получает конкретное измерение числом бит. Для корректного вычисления количества информации оказывается необходимым привлекать алгоритмическую теорию информации.

В машинном обучении (и в вычислительном индуктивном выводе вообще) такой подход был воплощен в нескольких концепциях. Среди этих концепций присутствуют такие, как *алгоритмическая вероятность* (АЛВ; Algorithmic Probability, ALP), разработанная Р. Соломоновым в 1964 г. [7]; принцип *минимальной длины сообщения* (МДС; Minimum Message Length, MML), предложенный К. Уалласом и Д. Болтоном в 1968 г. [8] (более поздние разработки по МДС можно найти, например, в работах [9, 10]), принцип *минимальной длины описания* (МДО; Minimum Description Length, MDL), описанный в 1978 г. Ж. Риссаненом [11] и несколько пересмотренный им позднее [12]; концепция идеальной МДО (Ideal MDL), предложенная М. Ли и П. Витани в 1989 г. [13] (см. также [14]), и *принцип аппроксимации сложности* (ПАС; Complexity Approximation Principle, CAP), введенный в работе [15]. Чуть ли не каждый из этих методов разными авторами связывается с формализацией бритвы Оккама [16, 17, 18, с. 10]

Хотя эти принципы несколько отличаются в деталях, основная идея у них одна, и ее можно сформулировать следующим образом (см., например, [19]). *Среди множества моделей следует выбрать ту, которая позволяет минимизировать сумму:* 1) длины описания модели (в битах); 2) длины данных, описанных посредством этой модели (в битах).

Общее правило, сформулированное таким образом, мы будем называть принципом минимальной длины описания (МДО). В случаях, когда речь будет идти о формальной те-

ории Риссанена, название которой и было выбрано для обозначения общего принципа, это будет оговариваться особо, чтобы не вызывать терминологической путаницы.

Длина описания модели определяет ее сложность, а длина описания данных с использованием модели — то, насколько хорошо модель удовлетворяет данным. Таким образом, компромисс между простотой и точностью модели приобретает объективный численный характер.

Прежде чем мы перейдем к более строгому рассмотрению вопроса сравнения гипотез с целью обоснования принципа МДО, а также к демонстрации различных его приложений, рассмотрим неформальное обоснование критерия простоты.

1.1.5. Бритва Оккама в научной эстетике и биологических системах

На удивление, два таких, казалось бы, разных вопроса, как «Что должно служить критерием истины?» и «Что такое красота?», оказываются тесно связанными через понятие простоты. Как уже отмечалось, простоту как критерий истинности связывают с принципом бритвы Оккама. Но и в эстетике (по крайней мере, научной, хотя есть основания думать, что не только в ней — см., например, работу [20] о применении понятия простоты в искусстве) критерий простоты находит свое применение. Так, всемирно известный математик Джордж Дейвид Биркгоф в 30-х годах XX века ввел уравнение красоты (см. [21–23], а также ряд других его работ, их переиздания и сборники) как

$$B = O/C,$$

где O — присутствующий в некотором объекте порядок (order); C — его сложность (complexity).

Примерно в это же время вышла в свет книга искусствоведа и драматурга В. М. Волькенштейна «Опыт современной эстетики», в которой, в частности, обосновывается идея того, что некоторый научный результат эстетичен, если с его помощью осуществляется сведение видимой сложности явлений к лежащей в их основе простоте. При этом также указывается, что, по замечаниям великих ученых, такое низведение сопряжено с сильными эстетическими переживаниями. И именно красота теории (естественно, при ее

согласованности с наблюдательными данными) часто служит критерием ее истинности, а слова «красивый», «прекрасный» нередко встречаются в трудах таких ученых, как Эйнштейн, Дарвин, Дирак, Больцман и многих других. Некоторые же из них, например Эйнштейн и Дирак, открыто говорили, что эстетический критерий является одним из важнейших в научном творчестве.

Несложно проследить, что выбор наиболее красивой теории в науке обычно оказывался и более успешным. Так, в астрономии развитие представлений о законах движения планет по небесной сфере, в частности переход от сложной системы эпициклов Птолемея к простой гелиоцентрической системе Коперника, можно охарактеризовать как выбор более красивых моделей. В химии таблица Д. И. Менделеева, безусловно, явилась красивым объяснением многообразия химических элементов. То же можно сказать и о дарвиновских законах развития в живой природе, и об открытии ДНК. В то же время все эти теории не только красивы, но и просты. Простота как значимый критерий красоты отмечается и в более поздних работах, например [24; 25].

Таким образом, успешность красивых теорий в науке может служить неформальным обоснованием критерия простоты, коль скоро простота и красота связаны. В связи с этим такие избитые выражения, как «краткость — сестра таланта» или «все гениальное просто», переосмысливаются и приобретают гораздо более глубокое значение.

Естественно, пока не существовало строгого определения понятия сложности, идеи Биркгофа и Волькенштейна (и, вероятно, многих других, здесь не упомянутых, людей) были не более чем уделом рефлексирующих ученых. Но в результате развития понятия алгоритмической сложности эти идеи нашли новый отклик [20, 26, 27], появилась также возможность их непосредственного применения в науке (см., например, [28, 29]), особенно в термодинамике и теории хаоса [30–32]. Интересно, что в рамках алгоритмической сложности понятие красоты естественным образом оказывается субъективным (основанным на личном опыте и на человеческой природе) [33].

Мы не можем более детально останавливаться на вопросе красоты в науке, так как не занимаемся здесь рассмотрением проблем ни научной методологии, ни эстетики, но не упомянуть о существовании тесной связи с принципом МДО было нельзя.

Другим источником неформального обоснования принципа минимальной длины описания являются данные, полученные в результате исследования принципов функционирования естественных нейронных сетей. Естественно, невозможно достаточно обоснованно утверждать, что работа человеческого мозга происходит согласно принципу МДО — об устройстве мозга еще слишком мало известно.

Тем не менее встречаются высказывания, согласно которым «очевидно, что биологические нейронные сети решили проблему бритвы Оккама» [6, с. 6], а сам принцип МДО используется в качестве критерия при выборе модели когнитивных процессов [34]. Это вызвано обоснованностью принципа МДО как подходящего критерия в индуктивном выводе, а обработка ощущений животными и человеком как раз и является индуктивным выводом, поэтому эта обработка с необходимостью должна следовать принципу МДО, чтобы быть корректной. Однако сейчас мы рассматриваем как раз обратный вопрос: существуют ли данные, подтверждающие, что естественные нейронные сети действительно следуют этому принципу?

Хотя мы не можем сделать такое заключение обо всем мозге в целом, есть данные, показывающие, что это верно, по крайней мере, для некоторых его подсистем. Эти результаты относятся преимущественно к системам первичной обработки сенсорной информации. Так, согласно Х. Барлоу [35–38], Д. Филду [39, 40] и ряду других авторов [41–44], важнейшей характеристикой обработки сенсорной информации в мозге является уменьшение ее избыточности или, иными словами, сжатие, что подтверждается различными исследованиями. Естественно, сжатие является не самоцелью, а результатом выделения из входного потока статистически независимых компонентов, порожденных различными источниками. Было бы интересно провести интерпретацию некоторых нейрофизиологических данных на основе принципа МДО, но, к сожалению, и этот вопрос выходит за рамки данной книги.

Помимо нейрофизиологических данных, подтверждающих сжатие информации на уровне отдельных групп нейронов, существуют и психологические исследования, показывающие значимость количественных информационных показателей для когнитивных процессов. Например, в работе [45] устанавливается тот факт, что скорость изучения человеком нового понятия строго зависит от алгоритмиче-

ской сложности этого понятия. К сожалению, как указано в работе [45], идея привлечения принципа МДО при исследовании человеческих когнитивных способностей лишь недавно попала в поле зрения ученых (см., например, [46–49]).

Таким образом, стремление к простоте с минимальной потерей информативности проявляется на разных уровнях: начиная от функционирования отдельных нейронов и заканчивая наукой в целом. Есть основания думать, что действие принципа МДО прослеживается и в других процессах. Например, ДНК, будучи своего рода моделью среды обитания, хотя и не может рассматриваться как минимальная программа, но определенная взаимосвязь между геномом и принципом минимальной длины описания в некоторых экспериментах прослеживается [50, 51]. Все это служит весьма сильным свидетельством в пользу данного принципа и заставляет искать причины такой универсальности. Далее в этой части книги мы попытаемся проследить логическую историю этого поиска, начиная с теоремы Байеса, а также обозначить все еще не решенные проблемы.

1.2. БАЙЕСОВСКИЕ МЕТОДЫ В ИНДУКТИВНОМ ВЫВОДЕ И МАШИННОМ ОБУЧЕНИИ

1.2.1. Теорема Байеса для выбора модели

Введем для начала некоторые определения, которые понадобятся нам для дальнейшего изложения.

Через выражение $\Pr(S)$ обозначим *вероятность* наступления некоторого события S в результате проведения испытания. В качестве такого события может выступать, например, «выпадение “решки”», а в качестве испытания — подбрасывание монетки.

Пусть задана *случайная величина* X, которая может принимать значения из некоторого множества $\mathrm{X} = \{x_1, x_2, \ldots, x_n, \ldots\}$. Например, X — это выпадающая в результате очередного подбрасывания сторона монетки; $\mathrm{X} = \{$«орел», «решка»$\}$. В результате единичного испытания случайная величина принимает одно и только одно значение.

Тогда *распределением вероятностей* случайной величины X называют отображение $P : \mathrm{X} \to [0, 1]$ такое, что $P(x_i) = \Pr(X = x_i)$, где выражение $\Pr(X = x_i)$ обозначает ве-