

УДК 616.89+371.128.1+371.3

СОЗДАНИЕ ОБУЧАЮЩЕЙСЯ СИСТЕМЫ – ЦЕНТРАЛЬНАЯ ПРОБЛЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

© А.В. Мелких

Melkikh A.V. Learning system design as the central problem of artificial intelligence. The article looks at the process of knowledge acquisition by the intellectual system and concludes that learning by the intellectual system is impossible.

ВВЕДЕНИЕ

Создание обучающейся системы, способной получать знания из окружающей среды, является центральной проблемой искусственного интеллекта. Существующие подходы к решению этой проблемы можно разделить на две группы [1]. К первой из них относятся исследования по математической психологии, ко второй – изучение поведения различного вида устройств (автоматов, машин, нейронных сетей и т. д.). Почему возникла такая проблема? При наблюдении за поведением человека и животных было замечено, что они имеют способность адаптироваться к окружающей среде. То есть при повторении некоторого внешнего сигнала реакция организма на него становится все более адекватной. Такой процесс называется обучением, и предполагается, что все организмы имеют *способность к обучению* определенного типа.

Однако оба указанных подхода к рассмотрению проблемы получения знаний (обучения) обладают существенными недостатками. С точки зрения математической психологии, под обучением понимается процесс приобретения всякого нового опыта организмом [1–4]. При этом предполагается, что система обладает способностью к обучению. Природа такой способности не обсуждается. В ряде работ моделируется приобретение знаний интеллектуальной системой. Однако при этом предполагается, что существуют программы, способные приобретать знания. Для этого, естественно, в программе должно быть заранее определено, что именно является знанием, а что нет.

«Техническое» направление имеет дело с описанием процесса поведения машин, все свойства которых точно определены [1, 5–16]. Однако при этом неявно предполагается, что кто-то задает свойства этих машин извне. Этот вопрос не рассматривается, а все доказательства, касающиеся дальнейшего поведения таких машин, исходят из того, что машины уже запрограммированы.

Вместе с тем, когда речь идет об обучении интеллектуальной системы (организма), то это может происходить только путем распознавания образов (то есть путем получения новой информации извне). Рассмотрим более подробно возможность обучения системы путем распознавания образов.

ЧТО ПОНИМАЕТСЯ ПОД ОБУЧЕНИЕМ НЕЙРОКОМПЬЮТЕРА?

Для моделирования способности организма к обучению предложены нейронные сети, которые, как предполагается, в основном обладают свойствами реальных нейронных систем. Однако под «обучением» нейросетей понимается нечто совсем иное, чем обучение организма!

Работа нейропроцессора начинается с предъявления образа. Предъявление образа (ввод первичного набора признаков) производится следующим образом (см., например, [5]). В начальный момент по *внешним связям* подаются сигналы, переводящие определенные элементы в активное состояние. Предъявленный образ поддерживается некоторое время, в течение которого связи «обучаются» (т. е. проводимость связей, по которым течет ток, уменьшается). После обучения процессор может распознавать экзаменуемые объекты, соотнося их с определенным классом из тех, которым он был обучен.

Все дело в том, что операция предъявления образа не может выполняться самопроизвольно и заключается в начальной фиксации связей между нейронами. Следовательно, нейросистема должна быть изначально приготовлена в определенном состоянии, и сама она этого сделать не может. Наиболее точное определение «обучения» сети дано в [7]: «разработка всевозможных алгоритмов построения синаптического связей, гарантирующих запоминание тех наборов образов, которыми почему-либо хотят заполнить память сети те или иные исследователи». Здесь слово «обучение» совершенно справедливо поставлено в кавычки, поскольку это ни что иное как задание начальных условий для системы. Именно исследователи задают этот набор образов, но не сам организм путем распознавания образов! Таким образом, мы имеем два существенно различных процесса: задание ключевых образов (начальных условий в нейронной сети) и накопление ценной информации, происходящее путем распознавания образов организмом.

Но если обучение существует, то в природе оно происходит именно путем распознавания образов! То есть рецепторы организма получают сигналы извне, в результате чего организм обучается. Все, чему организм

будет пытаться подражать, находится в окружающей среде. То есть вне системы. Для того чтобы прояснить физическую сущность этого процесса, рассмотрим обработку информации некоторой машиной (интеллектуальным роботом).

РАСПОЗНАВАНИЕ ОБРАЗОВ И ПРИНЯТИЕ РЕШЕНИЙ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМОЙ

Анализ различных определений понятия «обучение» показывает, что информация, получаемая системой в результате этого процесса, должна обладать следующими важными свойствами: 1) информация должна быть новой и 2) информация должна быть ценной для системы. Если хотя бы одно из этих свойств отсутствует, то обучения в таком процессе не происходит.

Итак, информация сначала регистрируется рецептором, потом происходит ее обработка (удаление шумов и т. д.) и распознавание образов. На последнем этапе машина должна принять решение. Рассмотрим, например, одну из простейших машин – машину Тьюринга [15, 17, 18].

Принцип Черча – Тьюринга утверждает, что любой физический процесс можно смоделировать на машине Тьюринга. Но машину Тьюринга обучить принципиально ничему нельзя! Можно лишь извне задать ей определенные правила поведения, алфавит. Принципиально, что сама машина правил для себя создать не может.

Если теперь перейти от машины Тьюринга к более сложным машинам, то легко видеть, что ситуация принципиально не изменится. Фон Нейман [14] говорит об эволюции автомата, как о запрограммированном процессе. То есть, чтобы получить сложное поведение, надо составить сложную программу («предусмотреть соответствующее логическое устройство»). Интеллектуальный робот должен хранить обобщенные модели объектов, которые в общем случае могут участвовать в ситуациях в среде [13, 16] и т. д.

Проведем мысленный эксперимент. Пусть в системе имеется M эталонов $A_1, \dots, A_M$. Пусть система встречается с новой ситуацией (новым образом), для которой не существует эталона из имеющегося набора. С одной стороны, система может просто не заметить этот образ и не предпринимать никаких действий (так же как, например, ультрафиолетовый свет не приводит ни к каким действиям при отсутствии соответствующих зрительных рецепторов). С другой стороны, система могла бы попытаться каким-нибудь образом обработать этот сигнал. Например, сделать его новым эталоном. Но как узнать, годится ли новый образ для этого? Допустим, что в системе существует алгоритм формирования новых эталонов. Этот алгоритм может быть реализован только на основе уже имеющихся эталонов:

$$A_{M+1} = F(A_1, \dots, A_M).$$

Но поскольку A_{M+1} не может быть эталоном, поскольку основным свойством эталонов является их независимость друг от друга. Такое свойство имеет место, например, в алфавите, используемом машиной Тьюринга. Кроме того, чтобы воспользоваться этим алгоритмом, мы обязательно должны предварительно

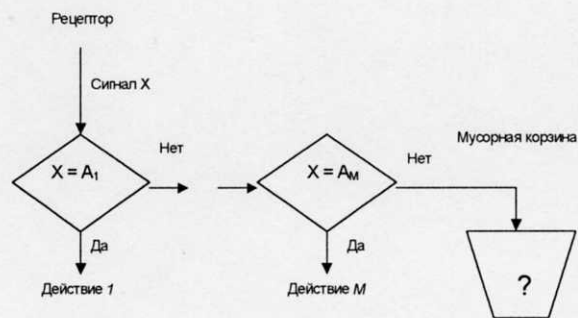


Рис. 1.

распознать образ и сделать вывод о том, к какому классу объектов он относится.

Можно попытаться просто сравнивать приходящие сигналы на предмет их повторяемости. Но что при этом будет происходить (рис. 1)?

Пусть имеется нераспознанный сигнал X_1 . Через некоторое время приходит сигнал X_2 . Что даст их сравнение между собой? Поскольку эталонов больше нет, то система не только не сможет определить, что для нее важнее: X_1 или X_2 , но и не сможет отделить любой из этих сигналов от шума! Таким образом, сколько бы сигналов система ни складывала в «мусорную корзину», никакого алгоритма, который мог бы делать что-нибудь с этими данными, у нее нет.

Анализ современных подходов к построению искусственного интеллекта [19] показывает, что эвристические программы и другие способы *всегда* подразумевают априорное задание тех свойств, которым машину необходимо обучить. Часто такое свойство программ рассматривается как их недостаток. Однако это не недостаток, а принципиальный запрет.

УСЛОВНЫЕ И БЕЗУСЛОВНЫЕ РЕФЛЕКСЫ

Рассмотрим с точки зрения физики формальных нейронов работу безусловного и условного рефлексов. В упрощенном виде работу безусловного рефлекса можно представить следующим образом: на рецептор поступает некоторый сигнал X , далее происходит сравнение этого сигнала с эталоном – A . Если сигнал совпадает с эталоном, то совершается определенное действие S_1 . Если сигнал не совпадает с эталоном, действие не совершается.

Теперь посмотрим, что собой представляет условный рефлекс (подробные схемы можно найти в литературе [20, 21]). Рецептор воспринимает два сигнала X и Y . Каждый из них проходит проверку на соответствие эталонам A и B . Далее, если идентифицированные сигналы A и B совпадают некоторое (заданное) количество раз K , то сигнал B уже в отсутствие сигнала A начинает вызывать действие S_2 .

Происходит ли при этом обучение системы? Константа K в счетчике совпадений является эталоном, который должен быть априорно задан. Далее, в системе должны существовать эталоны A и B , в противном случае она не сможет распознать оба эти сигнала. То есть в системе вся программа условного рефлекса (со всеми возможными результатами) уже априорно существует. В таком случае никакого обучения не произошло. Система просто выполнила определенное запрограммиро-

ванное действие. Понятно, что при этом необходимо выполнение определенных условий. Но ведь и безусловный рефлекс может быть сложным (комплексным), а значит, он будет содержать целый набор условий. Например, распознавание сложного образа происходит иерархически, т. е. путем последовательной проверки выполнения многих условий.

То есть условный рефлекс отличается от безусловного только внешне (по наблюдениям физиологов). Реализация того и другого вида рефлексов происходит на уровне элементарных процессов совершенно одинаково в том смысле, что и тот, и другой процессы являются детерминистскими (заранее запрограммированными), и при этом никакой новой ценной информации в системе не возникает.

Таким образом, если построение, приготовление, программирование системы обучением не называть (это более или менее очевидно, поскольку система не может этого сделать сама), то обучения не существует вообще! Легко видеть, что как только вы попытаетесь обучить систему определенным образом отвечать на внешний сигнал Z , эталона которого в ней нет, вы потерпите неудачу. Сколько бы раз этот внешний раздражитель ни действовал, система на него не реагирует.

В некотором смысле можно интеллектуальную систему считать черным ящиком, на вход которого подаются некоторые сигналы, а на выходе принимаются решения. Такой черный ящик независимо от того, как он устроен, должен иметь эталоны и программу работы с ними.

Что же на самом деле происходит, когда система встречается с чем-то новым? Происходит (при выполнении определенных условий) запуск априорных программ. Однако при наличии шума в системе какие-либо из программ могут не выполняться. При этом поведение системы окажется менее адекватным окружающей среде. Новые же полезные программы при этом не возникают. Вот почему организмы, будучи запрограммированы одинаково, могут вести себя по-разному.

ПОЧЕМУ НАМ КАЖЕТСЯ, ЧТО ЖИВОТНЫЕ ОБУЧАЮТСЯ?

Пусть в окружающей организм среде существуют одновременно сразу множество сигналов. Как оптимально организовать работу системы? С одной стороны, для каждого сигнала в организме имеется эталон (если его нет, то никакой адекватной реакции на сигнал так же нет). С другой стороны, организм не может заранее знать, какой именно образ окажется наиболее актуальным в данной ситуации. Причем, поскольку образов очень много, а эффекторов конечное количество (сравнительно небольшое), то может возникнуть ситуация, когда эффектор одновременно должен будет выполнить несколько различных (может быть, противоположных) функций, что невозможно (образов действительно может быть значительно больше, чем возможных действий с ними, но не наоборот). Тогда возникает задача выбора наиболее актуального действия. Пусть каждому образу соответствует своя полезная работа A_k . Необходимо построить стратегию распознавания и принятия решений так, чтобы суммарная полезная работа была максимальной. Конечно, это относится не к тем образам, появление которых приводит,



Рис. 2. Блок-схема программы

например, к немедленной гибели. Для их распознавания должна существовать безусловная схема, сразу запускающая априорные программы. Речь идет о «факкультативных» образах, полезная работа которых сравнительно невелика. При этом логично реагировать именно на те образы, которые вносят максимальный вклад в полезную работу (в условиях ограниченности времени и энергетических ресурсов эффекторов). Но априорно важность образов для данной ситуации неизвестна. Поэтому, по-видимому, наиболее выгодно не запускать эффекторы сразу, а дожидаться повторения образов. При этом выяснится, какой из них встречается чаще. Для этого необходимо в программу заложить определенное число таких повторений, потом сравнить полученные числа для разных образов и включить эффектор для тех из них, которые были повторены (за определенное время) максимальное количество раз.

Например, если есть последовательность:

ABCDAAAAACDAAAADCCCCAAA

то в ней программа, запускающая эффектор, соответствующий образу A, запустится (если число повторений равно, например, 10). В то же время на образ D, система не будет реагировать (при прочих равных условиях). Если, например, полезная работа, соответствующая образу C, достаточно велика, то он может запустить эффектор и при меньшем числе совпадений и т. д. В какой-то мере это экстраполяция начального временного кусочка описания среды на последующие моменты времени. Блок-схема программы представлена на рис. 2.

В конечном итоге все будет определяться балансом полезной работы и энергетических затрат всех эффекторов ΔU . Оптимальным будет вариант поведения, в котором затраты энергии минимальны.

Необходимо подчеркнуть, что в этом процессе все запрограммировано, и новый образ не может быть воспринят адекватно. Такая ситуация как раз и кажется нам обучением.

ЗАКЛЮЧЕНИЕ

Таким образом, анализ процессов рецепции информации, распознавания образов и принятия решений информационной системой показывает, что:

1. Обучение информационной системы невозможно, поскольку новая (нераспознанная) информация,

получаемая системой, не является ценной, а ценная информация (для нее есть соответствующие эталоны) не является новой.

2. Нет принципиальной разницы между условными и безусловными рефлексами. Нельзя выработать в организме незапрограммированный условный рефлекс. И в том, и в другом случае новой ценной информации организм не получает.

3. Поведение человека и животных полностью детерминировано. При возникновении в окружающей среде нового сигнала, в организме запускаются (или не запускаются) априорные программы.

4. При наличии шума (неконтролируемого внешнего влияния) часть программ в системе может не выполняться. При этом поведение системы станет менее адекватным. Новых программ при этом не возникает.

ЛИТЕРАТУРА

1. Приобретение знаний / Под ред. С. Осуги, Ю. Саэки. М.: Мир, 1990. 303 с.
2. *Аткинсон Р., Бауэр Г., Кротерс Э.* Введение в математическую теорию обучения. М.: Мир, 1969. 486 с.
3. *Осипов Г.С.* Приобретение знаний интеллектуальными системами. Основы теории и технологии. М.: Физматлит, 1997. 112 с.
4. *Венда В.Ф.* Системы гибридного интеллекта. Эволюция, психология, информатика. М.: Машиностроение, 1990. 448 с.
5. *Чернавский Д.С.* Проблема происхождения жизни и мышления с точки зрения современной физики // Успехи физических наук. 2000. Т. 170. № 2. С. 157-183.
6. *Grossberg S.* Studies of Mind and Brain. Boston: D. Riedel Publ. Co., 1982.
7. *Введенский В.Л., Ежов А.А.* Ритмы мозга и самовоспроизведение информации // Природа. 1990. №4. С. 33-44.
8. *Галушкин А.И.* Теория нейронных сетей // Нейрокомпьютеры и их применение. Кн. 1. М.: Издательское предприятие редакции журнала «Радиотехника», 2000. 416 с.
9. *Галушкин А.И.* Нейрокомпьютеры // Нейрокомпьютеры и их применение. Кн. 3. М.: Издательское предприятие редакции журнала «Радиотехника», 2000. 528 с.
10. *Круглов В.В., Борисов В.В.* Искусственные нейронные сети. Теория и практика. М.: Горячая линия «Телеком», 2001. 361 с.
11. *Хакен Г.* Информация и самоорганизация. Макроскопический подход к сложным системам. М.: Мир, 1991. 240 с.
12. *Малинецкий Г.Г., Потапов А.Б.* Современные проблемы нелинейной динамики. М.: Эдиториал УРСС, 2000. 336 с.
13. *Горский Н., Анисимов А., Горская Л.* Распознавание рукописного текста. От теории к практике. С.-Пб.: Политехника, 1997. 126 с.
14. *Хейман Дж. фон.* Теория самовоспроизводящихся автоматов. М.: Мир, 1971. 382 с.
15. *Turing A.M.* On computable numbers, with an application to the Entscheidungsproblem // Proceedings of LMS. 1936. Ser. 2. V. 42. № 3-4. P. 230-265.
16. *Каляев И.А., Гайдук А.Р.* Однородные нейрореподобные структуры в системах выбора действий интеллектуальных роботов. М.: «Янус-К», 2000. 280 с.
17. *Landauer R.* // IBM J. Res. Develop. 1961. V. 3. P. 183.
18. *Charles H. Bennett* // IBM J. Res. Develop. 1973. V. 17. P. 525.
19. *Andrew A.M.* Artificial intelligence, Viable Systems Chillaton, Devon (U.K.), Abacus Press, 1983.
20. *Гаазе-Раннопорт М.Г., Поспелов Д.А.* От амебы до робота: модели поведения, ПНТТЛ М.: Наука, 1987. 288 с.
21. *Блюм Ф., Лейзерсон А., Хофстердтор Л.* Мозг, разум и поведение. М.: Мир, 1988. 248 с.

БЛАГОДАРНОСТИ: Автор благодарит академика И.И. Еремина и профессора В.Д. Мазурова за поддержку идеи данной статьи.

Поступила в редакцию 26 ноября 2001 г.