

О ЗАДАЧЕ ОБОБЩЕНИЯ НАЧАЛЬНЫХ СИТУАЦИЙ

Сегодня существуют два обстоятельства, которые затрудняют как проведение работ по проблеме искусственного интеллекта, так — и в несравненно большей степени — понимание и оценку этих работ. Во-первых, в этой области до сих пор не сложилась иерархия целей и задач¹. Во-вторых, и это самое существенное, все такие работы либо предлагают лишь общие пути решения, либо дают грубые и приблизительные решения достаточно мелких по сравнению с проблемой в целом задач.

В этих условиях очень важно, решая какую-либо задачу, четко представлять себе и по возможности точно описывать, для каких целей ее решение может использоваться. Другими словами, надо знать структуру ее подзадач и ее надзадач. Только таким способом из нескольких неполных и неокончательных решений может быть выбрано одно — лишь зная, для чего нужна задача, можно сформулировать критерий качества ее решения.

Исходя из этих соображений, мы решили попробовать описать не только использование предложенного нами [1] языка в алгоритмах решения задачи обобщения ситуаций, но и в грубых чертах показать место этой задачи в одной из возможных иерархий подзадач, иерархии, начинающейся с самой верхней задачи, т. е. с конечной цели работ.

В первой части этой статьи формулируется задача прогноза и показывается, как при ее решении возникает необходимость обобщения ситуаций. Во второй части мы высказываем и обсуждаем одно требование к языку, предназначенному для решения задачи обобщения. Далее, вкратце излагается вариант языка, удовлетворяющий этому требованию. В третьей части обсуждаются операции обобщения записей, сделанных на предложенном языке. На примерах показывается, что, применяя эти операции в разумном порядке, можно получать хорошие обобщения на весьма малом числе примеров. Впрочем, вопросы, связанные с последовательностью применения операций, почти не обсуждаются.

1. ПОСТАНОВКА ЗАДАЧИ

Познание окружающего мира и выделение задачи прогноза. В качестве самой отдаленной, конечной цели моделирования мы возьмем одно из свойств человеческого интеллекта, а именно, его способность сознательно познавать природу. Эта цель еще слишком смутна и отдаленна. Как выделить наиболее важные задачи, кото-

¹ Другими словами, еще для очень многих задач неясно, на какие подзадачи их осмысленно разбивать и, наоборот, подзадачами каких проблем они являются.

рые нужно решить для достижения этой цели? В качестве ответа приведем слова физика Г. Герца [2]: «Ближайшая и в определенном смысле важнейшая задача нашего сознательного познания природы заключается в том, чтобы найти возможность предвидеть будущий опыт и в соответствии с этим регулировать наши действия в настоящем. Основой для решения этой задачи познания при всех обстоятельствах служит предшествующий опыт, полученный или из случайных наблюдений или из специальных экспериментов.

Метод, которым мы всегда пользуемся при выводе будущего из прошедшего, чтобы достигнуть этого предвидения, состоит в следующем: мы создаем себе внутренние образы или символы внешних предметов, причем создаем их такими, чтобы логически необходимые следствия этих представлений в свою очередь были образами естественно необходимых следствий отображенных предметов... Если нам удалось создать из накопленного опыта представления требуемого характера, то мы можем в короткое время вывести из них, как из моделей, следствия, которые сами по себе проявились бы во внешнем мире только через продолжительное время или же были результатом нашего вмешательства; следовательно, мы имеем возможность предвидеть факты и координировать принятые нами решения со сложившимися представлениями».

Итак, необходимо создать систему, способную прогнозировать результаты своих действий. *Прогнозом* для результата действия мы будем называть построение описания ситуации, в которой должна оказаться система после выполнения этого действия. Разумеется, прогноз будет, как правило, зависеть не только от действия, но и от ситуации, в которой находилась система к моменту начала действия.

Ясно, что система, способная прогнозировать, должна иметь органы чувств, показания которых она должна уметь записывать на внутреннем языке — создавать символические представления («внутренние образы») той обстановки, в которой она находится. Кроме того, такая система должна обладать памятью.

Описание памяти. Мы будем считать, что память заполняется записями вида $S_i \xrightarrow{D_i} A_i$, где S_i и A_i — некоторые описания ситуаций, а D_i — описание действия. Смысл каждой такой записи: в ситуации, которая удовлетворяла описанию S_i , было сделано действие D_i , и система оказалась после этого в ситуации, описываемой A_i . Будем называть такую запись *фактом*.

Все факты по их происхождению можно разбить на три группы. Во-первых, это записи реальных событий прошлого — экспери-

ментальные факты. В таких записях факт $S_i \xrightarrow{D_i} A_i$ интерпретируется так: S_i — описание ситуации (совокупность показаний рецепторов), в которой находилась система к началу действия D_i , а A_i — описание ситуации, в которой оказалась система к концу этого действия. Разумеется, S_i может включать в себя показания

рецепторов, относящиеся не только непосредственно к моменту начала действия, но и к предшествующим моментам, поскольку результат действия может зависеть также и от них.

Во-вторых, в памяти могут оказаться записи, которыми конструкторы ее загружают еще до того, как система приведена в действие — так сказать, «врожденные рецепты поведения»¹.

В-третьих, память заполняется записями — обобщениями накопленного опыта, которые формирует сама система. A_i в таких обобщениях может описывать лишь отдельные черты получающейся в результате действия ситуации, но и S_i будет содержать более мягкие требования, чем те опытные факты, на основании которых оно получено. В таких записях могут содержаться не только показания рецепторов, но и некоторые функции от них, выбираемые самой системой.

С помощью памяти такого рода можно планировать действия, необходимые для достижения какой-либо цели, и прогнозировать результат заданной последовательности действий.

Использование памяти для прогноза. Определение. Мы будем говорить, что ситуация M является *уточнением* ситуации L и записывать $M \rightarrow L$, если все требования, указанные в описании ситуации L , выполняются в ситуации M .

Это определение можно заменить на стандартное определение семантического следствия: будем говорить, что M есть уточнение L (L семантически следует из M), если в любом мире (модели), где выполняется M , выполняется и L . Однако до тех пор, пока не описаны язык и правила интерпретаций, этим достигается только кажущийся выигрыш в строгости.

Конечно, при построении системы необходимо заменить приведенные выше «определения» точным формальным определением, причем важно, чтобы по крайней мере в простых случаях проверка отношения уточнения производилась эффективным алгоритмом. Для языка, которым мы будем пользоваться в дальнейшем, такой алгоритм приведен в [1].

Теперь простейший вариант алгоритма прогноза результата действия можно описать так: пусть мы находимся в ситуации S_0 и нас интересует результат действия D_1 . Тогда находим в памяти такой факт $S_1 \xRightarrow{D_1} A_1$, что $S_0 \rightarrow S_1$ и считаем, что действие D_1 приводит к результату A_1 и в ситуации S_0 .

Если же нас интересует не результат действия D_1 , а результат последовательности действий $D_1 D_2$, то, найдя A_1 , мы не останавливаемся на этом, а находим факт $S_2 \xRightarrow{D_2} A_2$, такой, что $A_1 \rightarrow S_2$, и

¹ При отлаживании обучающихся систем в простых модельных средах необходимо следить за тем, чтобы объем априорных знаний был достаточно мал по сравнению со сложностью среды. Иначе по поведению системы невозможно будет оценить ее способность к обучению. В таких системах, как [3], и в какой-то степени [4], это требование не выполняется.

считаем, что A_2 — результат последовательности действий D_1D_2 . В принципе такая процедура действительно дает возможность предсказывать будущее на основе прошлого опыта.

Способ использования памяти для прогноза и сама структура памяти описаны здесь в очень грубых чертах. Мы не будем приводить здесь длинный список вопросов, которые необходимо разрешить, прежде чем модель памяти и прогноза можно будет использовать в действительно интересной системе. Заметим только, что даже в самой грубой модели должны вычисляться как сам прогноз, так и оценка вероятности его реализации (См. на эту тему [5].)

Описанный выше способ прогноза, даже если он и усовершенствован введением вероятностных оценок, основывается на следующей простой эвристике: если при удовлетворении некоторых условий определенное действие приводило к какому-то результату в прошлом, то и в будущем, при выполнении тех же условий, оно приведет к тому же результату. На важную роль этой эвристики во всем человеческом поведении указывал еще Д. Юм [6]. Мы будем называть ее *эвристикой I*. Можно считать, что в знакомых ситуациях естественность поведения системы будет обеспечиваться *эвристикой I*.

Но что же произойдет, если система окажется в новой для себя ситуации?

Задача прогноза в новой ситуации. Точно эту задачу можно сформулировать так: какой прогноз для действия D должна строить система, если она находится в такой ситуации S_0 , что в памяти не существует факта $S_1 \xrightarrow{D_1} A_1$ такого, что $S_0 \rightarrow S_1$ ¹? Другими словами, что делать, если описанная выше *эвристика I* не может быть применена?

Рассмотрим более частный случай. Пусть в памяти имеется много фактов вида $S_i \xrightarrow{D} A_i$ с тем действием D , результатом которого мы интересуемся. В этом случае может помочь следующая эвристика.

Эвристика II. Если количество разных A_i мало по сравнению с количеством фактов, т. е. действие D во многих разных случаях приводило к одному из небольшого числа A_i , то и в нынешней ситуации S_0 действие D тоже приведет к одному из этих уже встречавшихся A_i .

Казалось бы, такой случай нереален, поскольку почти всегда сделанные в разных ситуациях действия приведут к разным результатам, и количество разных A_i будет всегда близко к общему числу имеющихся в памяти фактов с действием D . Это в каком-то смысле верно, если считать, что A_i — это полное описание ситуа-

¹ В более совершенной модели это условие должно быть заменено на следующее: нет факта, который был бы способен дать прогноз, выполняющийся с высокой степенью вероятности.

ции, взятое непосредственно из опыта. Однако A_i может быть и неполным описанием ситуации, даже отдельным ее свойством. Тогда вполне вероятно, что число возможных A_i будет много меньше, чем общее число фактов в памяти.

Таким образом, использование эвристики II позволяет свести вопрос «к какому результату приведет действие D в ситуации S_0 ?» к совокупности вопросов вида «приведет ли действие D в ситуации S_0 к A_i (для всех i)?»

Далее мы будем считать, что такое сведение уже осуществлено. Точнее, мы будем говорить, что нам *задана задача прогноза* (D, A) , если задана пара: действие D и ситуация A . *Решить задачу прогноза* (D, A) в ситуации S_0 — значит ответить «да» или «нет» на вопрос «приведет ли в ситуации S_0 действие D к такому результату A' , что A' является уточнением A ?»

Задача прогноза и обобщение начальных ситуаций. Будем считать, что из памяти у нас отобраны а) множество фактов $S_i \xRightarrow{D} A_i$ таких, что $A_i \rightarrow A$, б) множество фактов $\tilde{S}_i \xRightarrow{D} \tilde{A}_i$ таких, что $\tilde{A}_i \not\rightarrow A$ и S_0 достаточно похоже хотя бы на одно S_i ¹.

Множество ситуаций $\{S_i\}$, состоящее из первых членов фактов, отобранных в п. а), назовем *таблицей присутствия* для задачи (D, A) [7].

Множество $\{\tilde{S}_i\}$ первых членов фактов из п. б) назовем *таблицей отсутствия*, а отдельные члены \tilde{S}_i этой таблицы — *противоречиями*.

Дадим еще одно определение. Любую мыслимую ситуацию L , такую, что, совершив в ней действие D , система перешла бы в такое M , что $M \rightarrow A$, мы назовем *положительной* относительно задачи прогноза (D, A) .

Таким образом, таблица присутствия — это некоторое множество положительных ситуаций, а именно множество тех положительных ситуаций, которые хранятся в памяти системы.

¹ Что значит, что одна ситуация похожа на другую? Даже в самой простой модели необходимо иметь не только алгоритмы, позволяющие устанавливать тот факт, что ситуация A является уточнением ситуации B [1], но и алгоритмы, которые могли бы, хотя и грубо, оценивать, насколько A похоже на B . В данном случае нам нужна симметричная функция $B(A, B)$, такая, что $B(A, B) = 0$, если и только если одновременно $A \rightarrow B$ и $B \rightarrow A$, причем $B(A, B)$ должно быть тем больше, чем более ситуации A и B не похожи друг на друга с интуитивной точки зрения. Конкретное построение этой функции зависит от языка, на котором описывается ситуация. Для языка, который мы будем употреблять, в качестве простейшего варианта функции B можно взять $B(A, B) = \rho(A, B) + \rho(B, A)$. Здесь ρ — функция, определенная в [1]. В принципе могут понадобиться различные модификации функции близости B , отличающиеся значениями на парах неравносмысленных выражений. Так, полезно иметь такую функцию, которая указывала бы, что разные показания одного рецептора больше «похожи» друг на друга, чем разные показания разных рецепторов. Такого рода изменения сводятся к изменению базиса в определении функции ρ в [1].

По условию $S_0 \notin \{S_i\}$. Задача состоит в том, чтобы решить, является ли S_0 положительной ситуацией. Для этого естественно попытаться построить описание всех вообще положительных ситуаций, чтобы проверить, выполняется ли оно для S_0 . Более точно, наша цель состоит в том, чтобы найти такое описание ситуации S , частными случаями которого являлись бы все положительные ситуации и только они.

Если мы нашли такое S , то в память можно записать факт $S \xRightarrow{D} A$. Теперь этим фактом можно пользоваться для прогноза, в частности, если $S_0 \rightarrow S$, то можно давать положительный ответ на задачу прогноза.

О п р е д е л е н и е. Задачу получения такого S , чтобы факт $S \xRightarrow{D} A$ был а) применим в возможно большем числе ситуаций и б) достаточно надежен, т. е. верен в большинстве ситуаций, к которым он применим, мы будем называть *задачей обобщения ситуаций*.

Будем полагать, что во всяком случае S должно удовлетворять следующим условиям:

- 1) для любого члена таблицы присутствия $S_i \rightarrow S$,
- 2) для любого противоречия $\bar{S}_i \nrightarrow S$.

Фактически в этом определении задан критерий качества задачи обобщения ситуаций как подзадачи в задаче прогноза. Сформулированный критерий, как и следует ожидать, зависит от мира, в котором будет существовать система. Но если мы заранее предназначаем систему для мира, похожего на человеческий, то такой критерий можно заменить на легко проверяемый, хотя и неформальный критерий естественности: хорошие алгоритмы обобщения, действуя в мире, подобном человеческому, должны порождать факты, кажущиеся естественными человеку — предсказывать «именно то, что предсказал бы я на его месте». Если это будет достигнуто, то можно надеяться, что система будет хорошо справляться с задачами прогноза, поскольку человек до сих пор довольно успешно решал такие задачи в своем мире.

2. ЗАДАЧА ОБОБЩЕНИЯ СИТУАЦИЙ И УТВЕРЖДЕНИЯ О СВЯЗЯХ

Итак, желательно построить описание всего множества положительных примеров, пользуясь лишь таблицами присутствия и отсутствия. Но если мы имеем в памяти не все положительные примеры, то где взять информацию, позволяющую описать всю их совокупность? На первый взгляд, недостающую информацию надо заложить в конструкцию алгоритма, который решает задачу обобщения. Алгоритм должен строить не все возможные для данной таблицы присутствия описания множества положительных примеров, а лишь их небольшую часть — естественное обобщение.

Однако информация, заложенная в алгоритм, будет одинакова для всех задач, решаемых этим алгоритмом. Вместе с тем в задачах

из разных областей множества положительных примеров могут быть построены по разным законам и требовать для своего описания разной информации. Поэтому, если мы хотим моделировать способность человека по ничтожному числу примеров успешно делать обобщения в задачах самого разного характера, то нам надо либо снабдить систему большим количеством разных специализированных алгоритмов¹, либо отказаться от надежды вложить всю необходимую информацию в конструкцию алгоритма. Последнее в самом деле возможно. Ведь можно считать, что у нас имеется двухступенчатый процесс. Вначале внешняя «реальная ситуация» преобразуется в ее представление на внутреннем языке, а затем эти представления преобразуются алгоритмом обобщения в обобщенную ситуацию. Дополнительная информация, необходимая для обобщения, может вводиться не только на второй стадии, но и на первой. Это значит, что в самих записях фактов будет содержаться информация о том, как устроено множество положительных ситуаций в данной задаче. Эта информация будет определяться и априорными эвристиками восприятия, и тем, что «знает» система об этой или аналогичной ей задаче. Таким образом, запись одной и той же внешней ситуации на внутреннем языке («восприятие» ситуации) должна зависеть от того, в каком состоянии находится система, какую задачу она решает. Теперь необходимо понять, какой язык будет способен легко и естественно выразить подобную «помогающую обобщению» информацию.

Какого рода информацию использует в процессе обобщения человек? Рассмотрим, например, работу физика-теоретика, который пытается на основе разрозненных экспериментальных фактов сформулировать общий закон природы. Можно отметить два наиболее неформализованных момента, в которых, собственно, и выражается опыт физика, используемый им в процессе обобщения. Во-первых, это умение выделить из всего множества параметров относительно небольшое число существенных. Во-вторых, после того, как отобраны существенные параметры, выносятся решение, связи каких параметров существенны в данной задаче. Другими словами, параметры разбиваются на небольшие группы так, чтобы в дальнейшем можно было рассматривать соотношения только между параметрами, принадлежащими одной группе. О параметрах, принадлежащих одной группе, говорят, что в данной задаче они связаны.

Приведем пример из истории механики, показывающий, насколько важны бывают утверждения о связях. Одним из серьезных препятствий развитию механики в средние века было мнение Аристотеля (кстати, полностью согласующееся с повседневным опытом), что в процессе движения существенна связь между скоростью в данный момент и силой, действующей на тело в тот же момент, так что «движущееся тело останавливается, если сила, его

¹ Одним из таких специализированных алгоритмов, где вся информация о задаче заложена в саму его конструкцию, является алгоритм, описанный в [8].

толкаящая, прекращает свое действие» (цит. по [9]). Революционная идея Ньютона, которая привела к бурному развитию механики и физики, заключалась в том, что он нашел параметры, связь между которыми действительно существенна в механике. «Основой классической механики, как она сформулирована Ньютоном, является связь между силой и изменением скорости, а не между силой и скоростью, как мы думали согласно интуиции» [9].

Заметим, что это не значит, что скорость стала несущественным параметром, поскольку предсказать, где будет тело через промежуток времени Δt , невозможно, не зная его скорости в данный момент; скорость как начальное условие определяет дальнейшее движение.

Приведем еще пример, из физиологии. В естественных движениях участвуют многие десятки мышц, работающих одновременно и согласованно. Однако для того, «чтобы высшие уровни центральной нервной системы могли эффективно решать задачи организации двигательных актов за требуемое время, необходимо, чтобы число управляемых параметров было не слишком большим» [10]. Поэтому в процессе обучения движению нервная система должна уметь, во-первых, выделить мышцы, управление которыми существенно в данном движении, и, во-вторых, определить, какие связи между управляющими воздействиями на мышцы существенны в данной задаче. Другими словами, для организации движения «осуществляется разбиение всех участвующих в движении мышц на небольшое число связанных групп» [10]. Здесь связь между мышцами означает, что между идущими к ним импульсами устанавливаются функциональные соотношения так, чтобы управляющие сигналы для всех мышц, связанных в группу, можно было бы выразить через один внешний параметр. Это и позволяет в процессе движения «управлять небольшим числом независимых параметров, хотя число мышц, участвующих в движении, может быть очень большим» [10].

Итак, мы видим, что как процесс обобщения, так и процесс обучения движению описывается в терминах разбиения переменных на связанные группы и выделения существенных переменных. В обоих случаях информация, специфичная для данной задачи и помогающая построить пригодное именно для нее обобщение (или управление движением), состоит в задании связей разной степени тесноты между параметрами. Поэтому в язык, которым будут пользоваться алгоритмы обобщения, в качестве элементарных утверждений нужно ввести утверждения о тесноте связей. Интересно отметить, что, поскольку на таком языке действия, необходимые для обобщения, описываются (на интуитивном уровне) так же, как и действия, необходимые для построения движений, можно надеяться, что алгоритм, построенный для обобщения, сможет решать задачи построения движений. В этом случае чем более сложные обобщения будут доступны системе, тем более сложным движениям сможет она обучиться. Напротив, богатый набор движений, совершаемых системой, будет означать, что ей доступны достаточно сложные обобщения.

Выделенное курсивом требование можно более формально высказать так. Язык, предназначенный для моделирования процессов обобщения, должен содержать в себе предикаты, определенные на выражениях самого языка, т. е. предикаты второго порядка (или средства, им эквивалентные), с помощью которых могут быть выражены утверждения о степени связи выражений языка.

Это требование находит некоторое подтверждение в работах по индуктивной логике. Дело в том, что задача, которую мы назвали «обобщение начальных ситуаций», практически совпадает с поставленной еще Ф. Бэконом [7] задачей индукции (настолько, что мы сочли возможным использовать его термины). Под этим названием она фигурирует в обширной литературе (см., например, [11] и приведенную там библиографию). Из нее можно сделать следующий вывод: для моделирования процессов обобщения недостаточно языка классической логики. Почти во всех работах к классической логике «прививаются» иные средства выражения. С формальной точки зрения эти средства позволяют строить утверждения о выражениях, записывающихся на языке классической логики. Другими словами, в язык, описывающий процесс индукции, вводятся предикаты, определенные на множестве формул основ-

ного языка. Эти предикаты используются для выражения того, насколько вероятна та или иная гипотеза при том или ином наборе опытных данных. Можно считать, что таким образом в каком-то смысле измеряется степень связи между гипотезой и данными опыта. Не отрицая важности подобных вероятностных связей, мы считаем, однако, что без формализации понятия связи того типа, который был описан нами выше, невозможно эффективно моделировать процессы обобщения.

Заметим, что поставленная нами задача обобщения ситуаций практически совпадает со стандартной постановкой задачи узнавания. Но в этой области существуют алгоритмы, которые хорошо (иногда не хуже человека) справляются с разнородными задачами довольно высокой сложности [12, 13]. Вместе с тем язык, который они используют, вовсе не содержит высказываний о связях. Как же выражается информация, описывающая специфику задачи? Дело в том, что фактически конечный результат работы такого алгоритма получается после суперпозиции двух процессов: процесса кодирования, который выполняется человеком, и собственно машинного алгоритма, который работает с уже закодированной информацией. При этом «кодирование» — не техническая работа, а сложный творческий процесс, требующий специальных знаний в той области, которой принадлежит решаемая задача. Именно в процессе «кодирования» человеком используется информация о связях, существенных в данной задаче. Таким образом, простота используемого в машинных алгоритмах языка, не содержащего высказываний о связях, приводит к тому, что значительная часть работы остается неформализованной и не поручается машине. Если же мы построим алгоритмы, которые будут явно использовать информацию о существенных связях, то можно надеяться, что специфика задачи будет отражаться не в интуитивном (и не контролируемом) способе «кодирования», а в информации о связях, которая должна быть в явном виде подана на вход алгоритма. Алгоритм же будет работать с имеющейся у него информацией совершенно единообразно. В результате мы сможем оценивать адекватность различных предположений о структуре связей и выбирать среди них наилучшие. В идеальном случае сам алгоритм должен не только использовать предоставленную ему информацию о связях, но и изменять ее, если имеющиеся факты (примеры, данные в обучении) плохо с ней согласуются.

Таким образом, мы считаем, что создание языка, в котором могут быть прямо выражены утверждения о связях, и алгоритмов, работающих с ним, интересно с точки зрения лучшего понимания того, что приводит к успеху или неудаче в решении какой-нибудь задачи узнавания.

В статье [1] мы попытались построить вариант языка, в котором формализовано понятие тесноты связей. С точки зрения поставленных здесь задач, т. е. описания мыслительных процессов, он может рассматриваться лишь как модель окончательного языка или, может быть, как его фрагмент. Однако нам кажется, что

в нем уже имеется ряд интересных для обсуждения свойств. Описание ситуаций в языке состоит из показаний рецепторов (элементарных высказываний о ситуации), соединенных связками \wedge и \vee . Эти связки отличаются от обычных тем, что они снабжены различными порядками. Порядок символизирует степень связи тех выражений, между которыми стоит связка — чем больше порядок, тем рыхлее связь. В свою очередь, степень связи имеет тот смысл, что поиск функциональных соотношений будет в первую очередь проводиться между тесно связанными рецепторами.

Теперь, когда выбран язык, встают два существенных вопроса. Во-первых, какими способами порождаются записи, несущие информацию о специфике задачи. Во-вторых, какими должны быть алгоритмы, использующие эту информацию для обобщения, в частности, для разбиения переменных на группы, внутри которых следует искать функциональные соотношения¹. Здесь мы лишь слегка затронем первый вопрос, после чего более подробно займемся вторым.

Мы будем считать, что некоторая часть системы обрабатывает непосредственно поступающую на входы информацию, перекодирует ее определенным образом (грубо говоря, вычисляет показания элементарных рецепторов для данной входной «картинки») и затем расставляет порядки связок. В эту часть системы должно быть с самого начала заложено, в частности, довольно большое число эвристик расстановки связок, в каком-то смысле «исходных принципов восприятия». Мы не будем здесь заниматься ими. В качестве примера можно дать следующие заведомо расплывчатые формулировки:

а) описания того, что соседствует во времени, связаны теснее, чем описания того, что разделено во времени;

б) описания того, что близко в пространстве, связаны теснее, чем описания того, что разнесено в пространстве;

в) описания свойств одного предмета связаны теснее, чем описания свойств разных предметов.

Кроме такого рода эвристик можно ввести априорные разделения на «сферы чувств», т. е. эвристики типа

г) все показания, относящиеся, например, к зрению, должны быть теснее связаны между собой, чем с показаниями, относящимися к положению руки.

Далее нам придется пользоваться сведениями о языке, изложенными в статье [1]. Здесь, в отличие от [1], мы будем для наглядности перед показаниями рецептора записывать и его название, так что запись $R\alpha$ будет означать, что показание рецептора R есть α .

¹ Хотелось бы также, чтобы такие алгоритмы не только использовали информацию, заложенную в структуре связок, но и меняли ее в процессе обобщения — делали более адекватной задаче.

Расстановку порядков связок в записи ситуаций иллюстрирует следующее выражение $T_4 \wedge_3 (B_x 1 \wedge_1 B_y 3) \wedge_3 ((P_X \wedge_1 P_M) \wedge_2 \wedge_2 (P_x 1 \wedge_1 P_y 0))$.

Первые два члена этой записи означают, что в момент времени 4 (T — рецептор времени) система находилась в точке с координатами (1, 3). Здесь B_x и B_y — рецепторы абсолютного положения системы. Последний (сложный) член записи кодирует состояние «руки» системы — показания ее проприорецепторов (рецепторов относительных координат руки P_x и P_y) и показания тактильного рецептора P . Запись $P_X \wedge_1 P_M$ означает, что в руке одновременно имеются и хлеб (X), и масло (M).

Возвращаясь к эвристикам расстановки связок, следует заметить, что система не сможет обойтись лишь заранее заложенными эвристиками такого рода. В зависимости от того, какая структура связок наиболее удобна в конкретной задаче, должны появляться новые способы расстановки связок, используемые, когда система занимается этой задачей. Или же применительно к задаче должны уточняться параметры в старых эвристиках (такие, как определение предмета в эвристике e). Мы не будем заниматься подробно алгоритмами появления новых эвристик расстановки связок, а в части 3 обсудим способы использования и изменения уже составленных связок.

3. ОПЕРАЦИИ И ПРИМЕРЫ ОБОБЩЕНИЯ

Наша цель состоит в том, чтобы найти преобразования, с помощью которых по таблице присутствия $\{S_i\}$ строится ситуация S , такая, что $S_i \rightarrow S$ для всех i (условие 1 на стр. 190), и при этом S выглядит достаточно естественно для человека. Разумеется, не надо забывать, что в результате этих преобразований не может получиться такое S , что для какого-нибудь j $\bar{S}_j \rightarrow S$, где \bar{S}_j — противоречие, т. е. член таблицы отсутствия (условие 2 на стр. 190). Эти преобразования должны формализовать, во-первых, отбрасывание несущественных параметров, и, во-вторых, разбиение параметров на группы, внутри которых только и проводится поиск функциональных связей.

Заметим, что поскольку смысл таблицы присутствия для задачи (D, A) состоит в том, что в любом из ее членов S_i действие D приводит к A , то вместо этого множества ситуаций мы можем рассматривать одну ситуацию, полученную связыванием связкой \vee всех членов таблицы присутствия. Если порядок этого \vee больше, чем ранг любого из соединяемых выражений, то, обозначив полученную таким способом ситуацию через Σ , можно утверждать, что условие $\Sigma \rightarrow S$ эквивалентно условию «для всех членов таблицы присутствия $S_i, S_i \rightarrow S$ ». Здесь (и в дальнейшем) \rightarrow означает формализацию отношения уточнения описания, данную в [1].

Итак, далее мы будем говорить об операциях, переводящих только что определенную ситуацию Σ (которую будем называть *общей ситуацией*) в обобщенную ситуацию S .

Выделение существенных параметров. Поскольку вопрос выделения существенных параметров сравнительно хорошо изучен, мы не будем уделять ему много внимания и рассмотрим лишь следующее элементарное правило.

Правило 1

Если в общей ситуации Σ есть подвыражение Q вида

$$Q \equiv (R\alpha_1 \bigvee_k R\alpha_2 \bigvee_k \dots \bigvee_k R\alpha_m),$$

причем m достаточно велико по сравнению с некоторым порогом, то подвыражение Q заменяется на 1, при условии, что результат этой замены S таков, что ни для одного противоречия \tilde{S}_j не выполняется $\tilde{S}_j \rightarrow S$ ¹.

Человек выдвигает гипотезу, что ситуация положительна для любого значения рецептора R (R — несущественный параметр), если из опыта известно, что для достаточно большого числа показаний рецептора R ситуация положительна. Правило 1 формализует выдвижение такого рода гипотез. Подобные гипотезы проверяются, исходя из имеющегося опыта. В системе этому соответствует проверка $\tilde{S}_j \rightarrow S$ для противоречий \tilde{S}_j .

Разумеется, преобразование типа «замена на 1», т. е. полное исключение показаний рецептора — это самая грубая форма обобщения. Выражение вида $(R\alpha_1 \bigvee_k R\alpha_2 \bigvee_k \dots \bigvee_k R\alpha_m)$ можно обобщать различными способами. Например, если область изменения α_i — отрезок, то можно вводить границы изменения α_i , т. е. фактически ограниченные кванторы; если значения R «разваливаются на несколько куч», то можно заменить во всех ситуациях рецептор R на новый, показания которого определяют лишь номер кучи [8], и т. д. В большей своей части преобразования такого типа скорее арифметические, чем логические. Однако намного более интересно вместо разработки методов обобщения в этом частном случае найти способы сведения к нему более широкого класса таблиц присутствия.

Правило 1 разумно использовать вместе с формальными преобразованиями выражений, описанными в [1]. Рассмотрим простейший пример, который показывает, как формальные преобразования расширяют область применимости правила 1.

Пример 1. Будем считать, что имеется два рецептора: рецептор времени T и рецептор «осязания» P . Показания последнего — это названия предметов, находящихся в данный момент в руке. Показание рецептора мы, как и выше, будем писать рядом с его названием, так что $T5$ означает момент времени 5, а PX

¹ Для любого выражения A в нашем языке (см. [1]) $A \bigwedge_k 1$ имеет тот же смысл, что и A , так что это преобразование просто исключает Q .

означает «хлеб в руке». Можно представлять себе, что выписанные ниже таблицы присутствия и отсутствия построены для задачи прогноза (D, A) , где D — действие «еда», а A — ощущение сытости.

Таблица присутствия

$$S_1 : T1 \underset{3}{\wedge} PX$$

$$S_2 : T4 \underset{3}{\wedge} PX$$

$$S_3 : T5 \underset{3}{\wedge} PX$$

Таблица отсутствия

$$\tilde{S}_1 : T2 \underset{3}{\wedge} P0$$

$$\tilde{S}_2 : T6 \underset{3}{\wedge} PK$$

$$\tilde{S}_3 : T8 \underset{3}{\wedge} PK$$

Нынешняя ситуация $S_0 : T9 \underset{3}{\wedge} PX$.

Содержательная интерпретация. В нескольких случаях (S_1, S_2, S_3) в разные моменты времени (T) в руке P находился хлеб X . Все это были «хорошие» ситуации. В других случаях, $\tilde{S}_1, \tilde{S}_2, \tilde{S}_3$, в другие моменты времени в руке либо вовсе ничего не было — 0, либо был камень — K . Это — противоречия, «плохие» ситуации. Нынешняя ситуация S_0 не совпадает ни с S_i , ни \tilde{S}_i .

В этом примере общая ситуация

$$\Sigma \equiv S_1 \underset{4}{\vee} S_2 \underset{4}{\vee} S_3 \equiv (T1 \underset{3}{\wedge} PX) \underset{4}{\vee} (T4 \underset{3}{\wedge} PX) \underset{4}{\vee} (T5 \underset{3}{\wedge} PX).$$

Поскольку в выражении Σ нет подвыражений вида $(R\alpha_1 \underset{k}{\vee} R\alpha_2 \underset{k}{\vee} \dots \underset{k}{\vee} R\alpha_m)$, то прямо к нему правило 1 применить нельзя.

Но если изменить форму выражения (не меняя смысла), то мы получим $\Sigma \sim PX \underset{3}{\wedge} (T1 \underset{4}{\vee} T4 \underset{4}{\vee} T5)$.

(Это следует из аксиомы 8 в [1]; \sim понимается здесь как равносмысленность, формализованная в нашем языке). После такого преобразования мы можем попробовать применить правило 1, заменяя $(T1 \underset{4}{\vee} T4 \underset{4}{\vee} T5)$ на 1. Полученное обобщение — $S \equiv PX \underset{3}{\wedge} \underset{3}{\wedge} 1 \sim PX$ (см. аксиому 10 в [10]), что означает «хлеб в руке». Правило 1 применимо, так как $T2 \underset{3}{\wedge} P0 \not\rightarrow PX$; $T6 \underset{3}{\wedge} PK \not\rightarrow PX$; $T8 \underset{3}{\wedge} PK \not\rightarrow PX$.

Разбиение параметров на группы. Попробуем теперь формализовать процесс разбиения параметров на группы, внутри которых имеются функциональные связи. Рассмотрим набор $\{\Pi_i\}$ всех положительных ситуаций для некоторой задачи прогноза (D, A) . Пусть в каждой ситуации Π_i выделено некоторое подвыражение π_i . Тогда набор $\{\pi_i\}$ не привязан к набору $\{\Pi_i\}$ в данной задаче, если для всех $j, k, F_{\pi_k}^{\pi_j} \lfloor \Pi_j \rfloor \in \{\Pi_i\}$, т. е. если резуль-

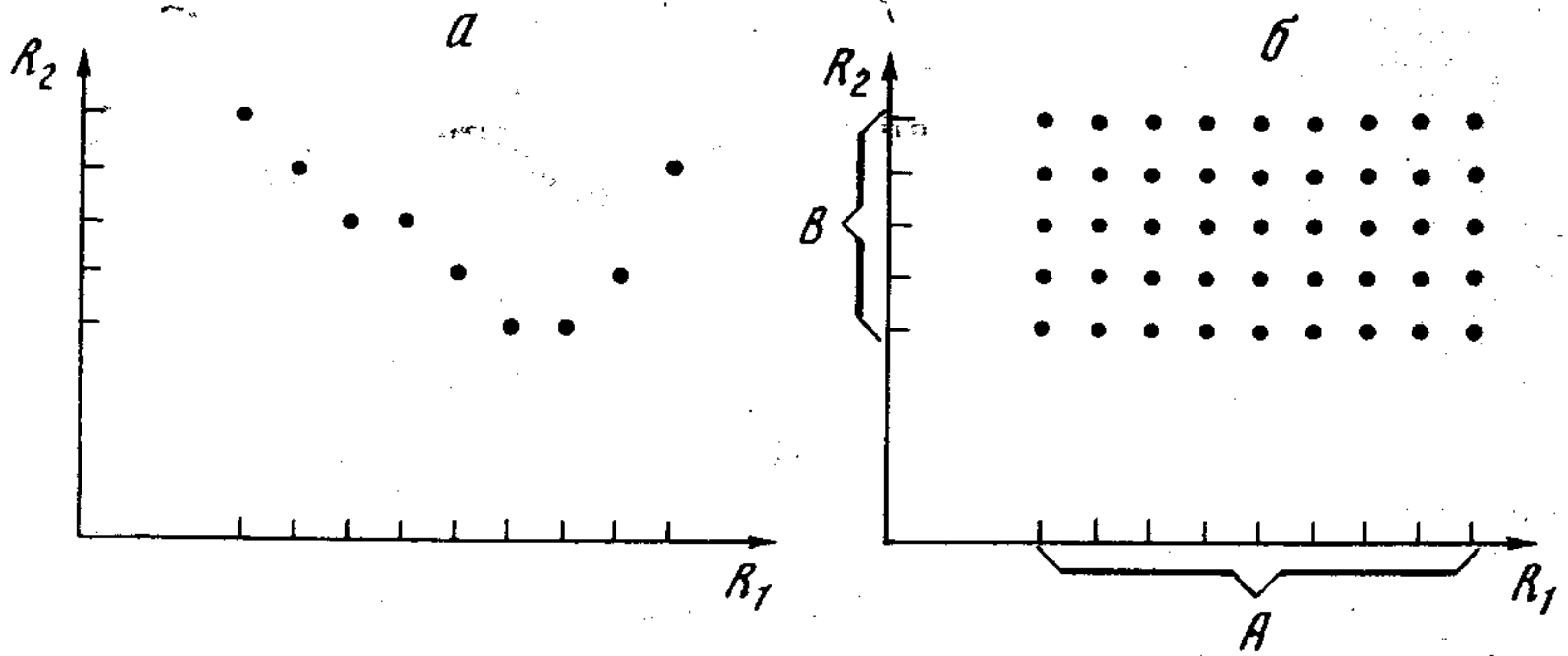


Рис. 1.
Множество положительных ситуаций (отмечены точками) в координатах R_1, R_2

а — набор значений рецептора R_1 привязан, **б** — не привязан к набору положительных ситуаций

тат замены в Π_j подвыражения π_j из ситуации Π_j на подвыражение π_k из ситуации Π_k тоже является положительной ситуацией.

Поясним это определение.

Случай 1. Рассмотрим следующий набор всех положительных ситуаций

$$\{\Pi_i\} = \{R_1\alpha_i \bigwedge_n R_2\beta_i\},$$

где $\beta_i = f(\alpha_i)$. В этом случае естественно сказать, что набор $\{R_2\beta_i\}$ привязан к набору $\{R_1\alpha_i \bigwedge_n R_2\beta_i\}$, так как между показаниями рецептора R_2 и остальной частью положительной ситуации имеется функциональная связь. Очевидно, если $R_1\alpha_i \bigwedge_n R_2\beta_i$ — положительная ситуация, то не существует такой положительной ситуации $R_1\alpha_j \bigwedge_n R_2\beta_j$, чтобы $\beta_j \neq \beta_i$ при том, что $\alpha_j = \alpha_i$ (рис. 1,а).

Случай 2. (Противоположный первому — между показаниями рецепторов в положительной ситуации нет никакой связи.) Набор всех положительных ситуаций имеет вид

$$\{\Pi_i\} = \{R_1\alpha_i \bigwedge_n R_2\beta_i\},$$

причем $\alpha_i \in A, \beta_i \in B$, где A и B — множества значений рецепторов R_1 и R_2 соответственно (рис. 1,б). Очевидно, в этом случае если $R_1\alpha_i$ встречалось в одной положительной ситуации, а $R_2\beta_j$ — в другой, то существует такая положительная ситуация $R_1\alpha_k \bigwedge_n R_2\beta_k$, что $\alpha_i = \alpha_k, \beta_j = \beta_k$. Именно в таких случаях по нашему определению набор $R_2\beta_i$ не привязан к $\{R_1\alpha_i \bigwedge_n R_2\beta_i\}$.

Нам кажется, что человек в своих рассуждениях очень часто и с большой охотой выдвигает гипотезы о том, что некоторый набор не привязан к другому. Если такая гипотеза не ведет к противоречию, то она принимается. Именно принятие таких гипотез и разбивает множество переменных на группы, только внутри которых существуют функциональные связи.

В нашей системе тоже можно проверить, не противоречит ли конкретная гипотеза такого типа экспериментальным фактам. Для этого рассматривается таблица присутствия, и из имеющихся в ней положительных ситуаций строятся ситуации, которые тоже должны быть положительными согласно проверяемой гипотезе. Если хотя бы для одного построенного таким образом выражения Π существует противоречие S_i , являющееся уточнением Π , то гипотеза противоречит фактам. Если же это не так, то гипотеза не противоречит фактам.

Хотелось бы иметь механизм, который строит естественные гипотезы о том, что один набор не привязан к другому. Заметим, что построение таких гипотез прямым перебором (с последующей проверкой на имеющихся фактах) приведет к тому, что значительная часть таких гипотез будет «предрассудками», т. е. будет оправдываться лишь для имеющихся в памяти положительных ситуаций, но не для всего их множества. Во избежание этого разумно воспользоваться тем, что известные соотношения внешнего мира уже в какой-то степени заложены в описание ситуаций расстановкой теснот связей. Это прежде всего закономерности, которые предвидел и заложил в способ записи воспринимаемых ситуаций конструктор. Кроме того, можно представить, что система сама изменяет структуру записи согласно раскрытым в прошлом опыте закономерностям мира.

Итак, имеется набор выражений S_i (или общая ситуация $\Sigma \equiv S_1 \underset{m}{\vee} S_2 \underset{m}{\vee} \dots \underset{m}{\vee} S_n$). В каждом S_i надо указать такое вхождение A_i , что набор $\{A_i\}$ не привязан к набору $\{S_i\}$. Различные подвыражения A_i отличаются, в частности, тем, что они привязаны к S_i связками разных порядков. Естественно считать, что если в каждом S_i мы выделили какое-нибудь A_i , привязанное самой «непрочной» связкой (т. е. связкой максимального в S_i порядка), то полученный набор $\{A_i\}$ имеет больше всего шансов быть не привязанным к набору $\{S_i\}$ в смысле нашего определения. (Если расстановка теснот связей правильно отражает существенные в данной задаче связи внешнего мира.)

Выделив такой набор A_1, A_2, \dots, A_n , мы можем, согласно этой гипотезе, заменить в каждом S_i выделенное нами вхождение A_i на любое A_k или сразу на выражение $A_1 \underset{m}{\vee} A_2 \underset{m}{\vee} \dots \underset{m}{\vee} A_n$

(обозначим результат такой замены через S'_i), а общую ситуацию Σ заменить на

$$\Sigma' \equiv S'_1 \underset{m}{\vee} S'_2 \underset{m}{\vee} \dots \underset{m}{\vee} S'_n.$$

Очевидно, Σ' будет обобщением, т. е. $\Sigma \rightarrow \Sigma'$. Если при этом нет противоречия \tilde{S}_i , такого, что $\tilde{S}_i \rightarrow \Sigma'$ (обобщение не противоречит экспериментальным фактам), то гипотезу о непривязанности набора $\{A_i\}$ можно было бы считать приемлемой. Однако примеры показывают, что условия, наложенные выше на выбор A_i , еще не обеспечивают естественности гипотезы, т. е. полученное из Σ выражение Σ' часто выглядит крайне неестественно. Из анализа примеров вытекает, однако, что Σ' выглядит вполне естественно, если выбранные A_i похожи друг на друга. (Здесь нам снова нужна мера близости выражений, о которой говорилось в сноске на стр. 189). Поэтому мы выдвигаем требование, чтобы все A_i были близки друг к другу. В этом случае Σ' можно считать «естественным» обобщением, хотя быть может еще недостаточным для наших целей.

Правило 2. Если в выражении $S_1 \underset{m}{\vee} S_2 \underset{m}{\vee} \dots \underset{m}{\vee} S_n$ подвыражения A_i выражения S_i близки друг к другу и, кроме того, привязаны к S_i связкой максимального в S_i порядка, то выражение $S_1 \underset{m}{\vee} S_2 \underset{m}{\vee} \dots \underset{m}{\vee} S_n$ заменяется на выражение $S'_1 \underset{m}{\vee} S'_2 \underset{m}{\vee} \dots \underset{m}{\vee} S'_n$, где S'_i — результат замены в выражении S_i выделенного вхождения A_i на выражение $A_1 \underset{m}{\vee} A_2 \underset{m}{\vee} \dots \underset{m}{\vee} A_n$, при условии, что не существует такого противоречия \tilde{S}_j , что $\tilde{S}_j \rightarrow S'_1 \underset{m}{\vee} S'_2 \underset{m}{\vee} \dots \underset{m}{\vee} S'_n$.

Это правило носит слишком общий характер и ниже мы будем применять лишь следующий его частный вариант: если S_i имеет вид $A_i \underset{k}{\wedge} B_i$, где $\underset{k}{\wedge}$ — связка максимального в S_i порядка, то

$(A_1 \underset{k}{\wedge} B_1) \underset{m}{\vee} (A_2 \underset{k}{\wedge} B_2) \underset{m}{\vee} \dots \underset{m}{\vee} (A_n \underset{k}{\wedge} B_n)$ заменяется на

$$S \equiv ((A_1 \underset{m}{\vee} A_2 \underset{m}{\vee} \dots \underset{m}{\vee} A_n) \underset{k}{\wedge} B_1) \underset{m}{\vee} \dots \underset{m}{\vee} ((A_1 \underset{m}{\vee} A_2 \underset{m}{\vee} \dots \dots \underset{m}{\vee} A_n) \underset{k}{\wedge} B_n) \sim (A_1 \underset{m}{\vee} A_2 \underset{m}{\vee} \dots \underset{m}{\vee} A_n) \underset{k}{\wedge} (B_1 \underset{m}{\vee} B_2 \underset{m}{\vee} \dots \underset{m}{\vee} B_n).$$

Последний переход справедлив, когда m больше, чем максимальный порядок связок во всех A_i (см. [1], аксиома 8). Это будет автоматически выполняться во всех наших примерах.

Видно, что это преобразование не просто одно из возможных обобщений; оно действительно разбивает параметры на отдельные не связанные группы. Если в первоначальном выражении в принципе можно было искать функциональную связь между параметрами, входящими в A_i и в B_i , то в S «переменные разделились», и ясно, что функциональную зависимость можно искать только внутри B_i и отдельно внутри A_i . Рассмотрим очень простой пример.

Пример 2. Отличие от примера 1 только в том, что в ситуации S_1 «в руке» был не только «хлеб», но и «масло» (M), т. е.
 $S_1 : T1 \underset{3}{\wedge} (PX \underset{1}{\wedge} PM)$.

Тогда общая ситуация

$$\Sigma \equiv (T1 \underset{3}{\wedge} (PX \underset{1}{\wedge} PM)) \underset{4}{\vee} (T4 \underset{3}{\wedge} PX) \underset{4}{\vee} (T5 \underset{3}{\wedge} PX).$$

Решение

1. Применяем правило 2, считая, что $A_1 \equiv T1$, $A_2 \equiv T4$, $A_3 \equiv T5$.

Получаем

$$\Sigma_1 \equiv (T1 \underset{4}{\vee} T4 \underset{4}{\vee} T5) \underset{3}{\wedge} ((PX \underset{1}{\wedge} PM) \underset{4}{\vee} PX \underset{4}{\vee} PX).$$

Вынесение PX за скобки во втором члене и дальнейшие преобразования дают

$$\begin{aligned} \Sigma_1 &\sim (T1 \underset{4}{\vee} T4 \underset{4}{\vee} T5) \underset{3}{\wedge} (PX \underset{1}{\wedge} (PM \underset{4}{\vee} 1 \underset{4}{\vee} 1)) \sim \\ &\sim (T1 \underset{4}{\vee} T4 \underset{4}{\vee} T5) \underset{3}{\wedge} (PX \underset{1}{\wedge} 1) \sim (T1 \underset{4}{\vee} T4 \underset{4}{\vee} T5) \underset{3}{\wedge} PX. \end{aligned}$$

Заметим, что если бы в качестве A_i мы выбрали бы не «похожие» выражения $T1$, $T4$, $T5$, а такие разнородные, как $T1$, $T4$ и PX , то мы получили бы весьма неестественное выражение

$$(T1 \underset{4}{\vee} T4 \underset{4}{\vee} PX) \underset{3}{\wedge} ((PX \underset{1}{\wedge} PM) \underset{4}{\vee} PX \underset{4}{\vee} T5) \sim (T1 \underset{4}{\vee} T4 \underset{4}{\vee} \underset{4}{\vee} PX) \underset{3}{\wedge} (PX \underset{4}{\vee} T5).$$

2. Применяя к первой скобке правило 1, получаем окончательную обобщенную ситуацию $S = 1 \underset{3}{\wedge} PX \sim PX$.

В данном примере это означает, что для того, чтобы после выполнения действия «еда» возникло ощущение сытости (прогнозируемое A), нужно иметь в руке по крайней мере «хлеб».

Пример 3

Таблица присутствия

$$S_1 : T1 \underset{3}{\wedge} (B_x1 \underset{1}{\wedge} B_y3) \underset{3}{\wedge} (PX \underset{2}{\wedge} (P_x0 \underset{1}{\wedge} P_y0)),$$

$$S_2 : T3 \underset{3}{\wedge} (B_x4 \underset{1}{\wedge} B_y1) \underset{3}{\wedge} (PX \underset{2}{\wedge} (P_x1 \underset{1}{\wedge} P_y0)),$$

$$S_3 : T4 \underset{3}{\wedge} (B_x1 \underset{1}{\wedge} B_y1) \underset{3}{\wedge} (PX \underset{2}{\wedge} (P_x1 \underset{1}{\wedge} P_y1)),$$

$$S_4 : T5 \underset{3}{\wedge} (B_x2 \underset{1}{\wedge} B_y2) \underset{3}{\wedge} (PX \underset{2}{\wedge} (P_x0 \underset{1}{\wedge} P_y1)).$$

Таблица отсутствия

$$\tilde{S}_1 : T2 \underset{3}{\wedge} (B_x2 \underset{1}{\wedge} B_y2) \underset{3}{\wedge} (P0 \underset{2}{\wedge} (P_x1 \underset{1}{\wedge} P_y1)),$$

$$\tilde{S}_2 : T6 \underset{3}{\wedge} (B_x4 \underset{1}{\wedge} B_y3) \underset{3}{\wedge} (PX \underset{2}{\wedge} (P_x0 \underset{1}{\wedge} P_y1)),$$

$$\tilde{S}_3 : T7 \underset{3}{\wedge} (B_x5 \underset{1}{\wedge} B_y2) \underset{3}{\wedge} (P0 \underset{2}{\wedge} (P_x0 \underset{1}{\wedge} P_y0)).$$

Нынешняя ситуация $S_0: T8 \wedge (B_x3 \wedge B_y3) \wedge (PX \wedge (P_x1 \wedge P_y0))$.

Содержательная интерпретация. Здесь T и P означают, как и раньше, рецепторы времени и «осязания», P_x и P_y — это относительные координаты положения руки (будем считать, что рука может находиться всего в 4 положениях относительно тела, так что P_x и P_y принимают значения 0 и 1), а B_x и B_y — координаты самой системы (рис. 2). В данном случае для того, чтобы действие «еда» приводило к результату «сытость»,

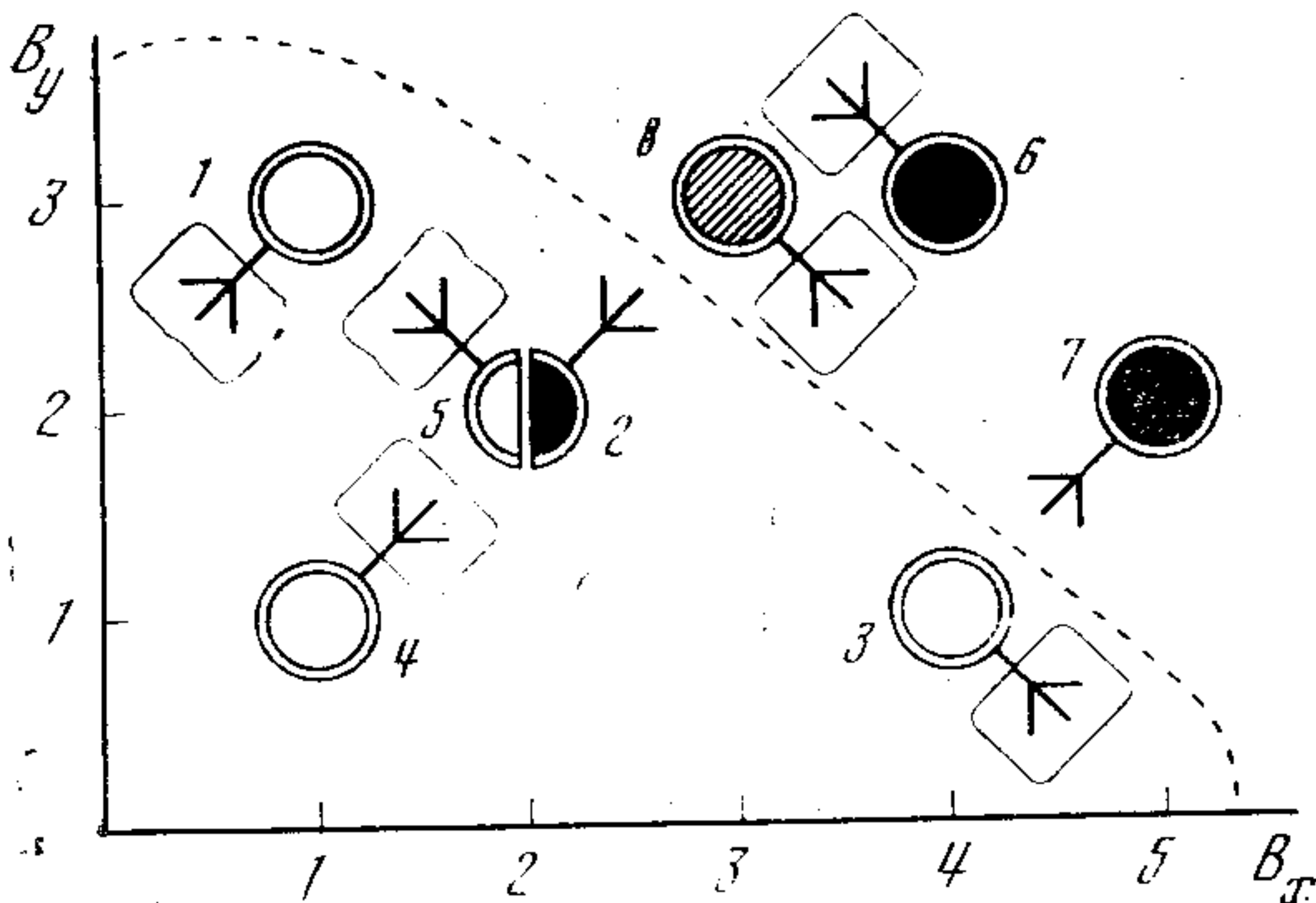


Рис. 2.
Пример 3

Положение системы в разные моменты времени (цифры рядом с кружками) в пространстве. Белыми кружками обозначены положительные ситуации, черными — противоречия, штриховкой — нынешняя ситуация

необходимо, чтобы, во-первых, в руке был хлеб и, во-вторых, чтобы сама система находилась в «столовой» — некоторой области пространства, обозначенной на рисунке пунктирной границей. Конечно, система ничего не знает об этой границе, ей известны только положительные и отрицательные ситуации. Отрицательными здесь называются ситуации, где либо в руке не было хлеба, либо система находилась вне столовой.

Ясно, что в этом примере запись адекватна рассматриваемой задаче, так как исходная закономерность состоит, в частности, в связи координат B_x и B_y (границы «столовой» описываются выражением вида $f(B_x, B_y) = 0$, а в записях B_x и B_y связаны самой тесной связкой и слабо привязаны к остальному).

Решение

Мы не будем подробно выписывать получающиеся громоздкие выражения. Ряды одинаково устроенных членов будут с некоторого момента заменяться многоточиями.

Общая ситуация

$$\Sigma \equiv (T1 \wedge (B_x1 \wedge B_y3) \wedge (PX \wedge (P_x0 \wedge P_y0))) \vee \vee (T3 \wedge (B_x4 \wedge B_y1) \wedge (PX \wedge (P_x1 \wedge P_y0))) \vee \dots$$

1. Применяя правило 2, можно отделить время T от описаний ситуаций

$$\Sigma_1 \equiv (T1 \vee T3 \vee T4 \vee T5) \wedge (((B_x1 \wedge B_y3) \wedge (PX \wedge (P_x0 \wedge P_y0))) \vee \dots)$$

2. Применяя к первому конъюнктивному члену Σ_1 правило 1, можно вообще избавиться от упоминания времени в описании ситуации

$$\Sigma_2 \equiv ((B_x1 \wedge_1 B_y3) \wedge_3 (PX \wedge_2 (P_x0 \wedge_1 P_y0))) \vee_4 ((B_x4 \wedge_3 B_y1) \wedge_3 \wedge_3 (PX \wedge_2 (P_x1 \wedge_1 P_y0))) \vee_4 \dots$$

3. Снова применяем правило 2 к полученной ситуации и имеем

$$\Sigma_3 \equiv ((B_x1 \wedge_1 B_y3) \vee_4 (B_x4 \wedge_3 B_y1) \vee_4 \dots) \wedge_3 ((PX \wedge_2 (P_x0 \wedge_1 \wedge_1 P_y0)) \vee_4 \dots).$$

4. Теперь попытка применить правило 2 к первому конъюнктивному члену дает ситуацию

$$\Sigma_4 \equiv ((B_x1 \vee_4 B_x4 \vee_4 \dots) \wedge_1 (B_y3 \vee_4 B_y1 \vee_4 \dots)) \wedge_3 ((PX \wedge_2 \wedge_2 (P_x0 \wedge_1 P_y0)) \vee_4 \dots).$$

Видно, что она не может служить обобщением ситуаций таблицы присутствия, поскольку одно из противоречий (\tilde{S}_2) становится ее частным случаем $\tilde{S}_2 \rightarrow \Sigma_4$. Правило 2 неприменимо.

Обратимся поэтому ко второму члену выражения Σ_3 , а именно

$$(PX \wedge_2 (P_x0 \wedge_1 P_y0)) \vee_4 (PX \wedge_2 (P_x1 \wedge_1 P_y0)) \vee_4 \dots$$

Вынесение за скобки PX дает

$$PX \wedge_2 ((P_x0 \wedge_1 P_y0) \vee_4 (P_x1 \wedge_1 P_y0) \vee_4 (P_x1 \wedge_1 P_y1) \vee_4 (P_x0 \wedge_1 P_y1)).$$

Повторное применение правила 2 преобразует это выражение в

$$PX \wedge_2 ((P_x0 \vee_4 P_x1 \vee_4 P_x1 \vee_4 P_x0) \wedge_1 (P_y0 \vee_4 P_y0 \vee_4 P_y1 \vee_4 P_y1)).$$

5. Поскольку координаты руки P_x и P_y могут принимать только два значения, к каждой из последних скобок применяется правило 1; в результате весь член превращается в PX , а обобщенная ситуация есть

$$S \equiv PX \wedge_3 ((B_x1 \wedge_1 B_y3) \vee_4 (B_x4 \wedge_1 B_y1) \vee_4 (B_x1 \wedge_1 B_y1) \vee_4 \vee_4 (B_x2 \wedge_1 B_y2)).$$

Заметим следующее: а) $S_0 \not\rightarrow S$, т. е. в данном случае мы не можем прогнозировать «сытость», что соответствует содержательной интерпретации, б) полученное обобщение весьма естественно; вероятно, и человек пришел бы к аналогичному заключению: во-первых, из положительных примеров видно, что хлеб, по-видимому, существует, а, во-вторых, противоречие \tilde{S}_2 показывает,

что дело не только в хлебе, а вероятно еще и в координатах в) если рассматривать результат как окончательный, то надо заметить, что нами получено чрезвычайно слабое обобщение. Вместо того чтобы описать множество точек «столовой» или хотя бы высказать о нем правдоподобные гипотезы, мы пришли к выводу, что выполнять действие «еда» имеет смысл лишь в тех точках пространства, где мы уже бывали «сытыми». Логично, но неразумно. Однако с точки зрения выделения параметров, между которыми надо искать функциональные связи, мы добились полного успеха. Если в исходном выражении мы имели шесть параметров, между которыми могли искать функциональные связи, то конечное выражение ясно показывает, что возможна лишь связь между координатами B_x и B_y , что полностью соответствует содержательному смыслу задачи¹.

Теперь при помощи других методов надо построить функцию от B_x и B_y , аппроксимирующую форму «столовой». Эта задача несравненно проще. Это упрощение удалось осуществить потому, что была использована информация, выраженная в порядках связей. В данном случае расстановка связей была адекватна задаче — тесно связанными между собой оказались те переменные, между которыми с содержательной точки зрения в данной задаче существенна функциональная связь. Действительно ли этот результат был получен с использованием информации о связях или просто положительные примеры и противоречия были так подобраны, что и без этой информации можно было прийти к такому же решению? Чтобы выяснить это, можно рассмотреть следующий пример.

Пример 4. В этом примере положительные ситуации и противоречия отличаются от примера 3 лишь тем, что в них опущена информация о связях, а именно все связки заменены на связку \bigwedge_1 .

Не выписывая решения, скажем только, что применениями правил 1 и 2 можно, в частности, в качестве обобщенной ситуации S получить не только ту, которую мы уже имели в решении примера 3

$$PX \bigwedge_3 ((B_x1 \bigwedge_1 B_y3) \bigvee_4 (B_x4 \bigwedge_1 B_y1) \bigvee_4 (B_x1 \bigwedge_1 B_y1) \bigvee_4 (B_x2 \bigwedge_1 B_y2)),$$

но и, например, такую:

$$(B_x1 \bigwedge_1 P_x0) \bigvee_2 (B_x4 \bigwedge_1 P_x1) \bigvee_2 (B_x1 \bigwedge_1 P_x1) \bigvee_2 (B_x2 \bigwedge_1 P_x0),$$

в зависимости от последовательности применения этих правил. Последнее выражение крайне неестественно и непригодно для дальнейшей работы. К тому же оно будет давать неверные прог-

¹ Показание PX хотя и существенно само по себе, но привязано к координатам связкой значительно более рыхлой, чем та, что связывает их между собой, поэтому в первую очередь надо искать соотношение между B_x и B_y .

нозы. В примере 3 оно не могло быть получено именно благодаря использованию порядков связок.

Изменение структуры записи ситуаций. Итак, если структура связок адекватна раскрываемой закономерности, то наши правила, использующие информацию о связях, направляют перебор и в известной степени избавляют от предрассудков и неестественных обобщений. Но что делать, если структура выражения не отражает искомый закон? Здесь надо заметить, что из примера 4 не следует делать вывод, что любое нарушение структуры выражений приводит к предрассудкам. Если противоречия мало отличаются от положительных примеров (это значит, что они отличаются только в существенном!), то с их помощью можно отбросить все предрассудки. (Именно это требование, а не само по себе количество противоречий и положительных примеров, является существенным фактором для оценки качества наличного экспериментального материала.) Поэтому (казалось бы, в противоречии с предыдущей идеологией) мы предлагаем следующее. Если таблица присутствия такова, что правило 2 неприменимо из-за противоречий (это и означает неадекватность теснот связей искомому закону), то ко всем членам таблицы присутствия применяется правило 3.

Правило 3.

Если n — максимальный порядок связок типа \bigwedge , встречающихся в Q , а k — второй по величине такой порядок, то перевести Q в Q' , где Q' отличается от Q тем, что все вхождения \bigwedge_k заменены на вхождения \bigwedge_n , при условии, что не существует противоречия \tilde{S}_j , такого, что $S_j \rightarrow Q'$.

Приведем пример, иллюстрирующий применение этого правила.

Пример 5.

Таблица присутствия

$$S_1 : (B_x 1 \bigwedge_1 B_y 1) \bigwedge_2 (P_x 1 \bigwedge_1 P_y 1),$$

$$S_2 : (B_x 2 \bigwedge_1 B_y 2) \bigwedge_2 (P_x 0 \bigwedge_1 P_y 0),$$

$$S_3 : (B_x 2 \bigwedge_1 B_y 4) \bigwedge_2 (P_x 0 \bigwedge_1 P_y 1),$$

$$S_4 : (B_x 1 \bigwedge_1 B_y 4) \bigwedge_2 (P_x 1 \bigwedge_1 P_y 1).$$

Таблица отсутствия

$$\tilde{S}_1 : (B_x 2 \bigwedge_1 B_y 1) \bigwedge_2 (P_x 1 \bigwedge_1 P_y 0),$$

$$\tilde{S}_2 : (B_x 1 \bigwedge_1 B_y 4) \bigwedge_2 (P_x 0 \bigwedge_1 P_y 1).$$

Нынешняя ситуация $S_0 : (B_x 1 \bigwedge_1 B_y 3) \bigwedge_2 (P_x 1 \bigwedge_1 P_y 0).$

Содержательная интерпретация. На полосе поля шириной в одну клетку (между значениями $B_x 1$ и $B_x 2$) растет трава, и нигде больше травы нет. Система в нескольких случаях находилась в точках поля, соседних с травой, и рука при этом находилась над клеткой с травой (это положительные ситуации). В других случаях рука была повернута в сторону от травы — это противоречия. Впрочем, как это видно из записи ситуаций, самой системе информация о местонахождении травы недоступна, и ей приходится строить обобщение, используя лишь сведения об относительном положении руки и своем местонахождении. Общая картина изображена на рис. 3.

Следует отметить, что в этой задаче, несмотря на небольшое количество материала, удастся построить правильное обобщение, раскрывающее имеющуюся закономерность («травя под рукой») так, что даже ситуацию S_0 , существенно отличную от положительных примеров (новая точка пространства; относительные координаты руки, встречавшиеся ранее только в ситуации \tilde{S}_1 — противоречии), удастся включить как частный случай в построенное обобщение.

Решение. Общая ситуация

$$\Sigma \equiv ((B_x1 \wedge B_y1) \wedge (P_x1 \wedge P_y1)) \vee ((B_x2 \wedge B_y2) \wedge (P_x0 \wedge P_y0)) \vee \\ \vee ((B_x2 \wedge B_y4) \wedge (P_x0 \wedge P_y1)) \vee ((B_x1 \wedge B_y4) \wedge (P_x1 \wedge P_y1)).$$

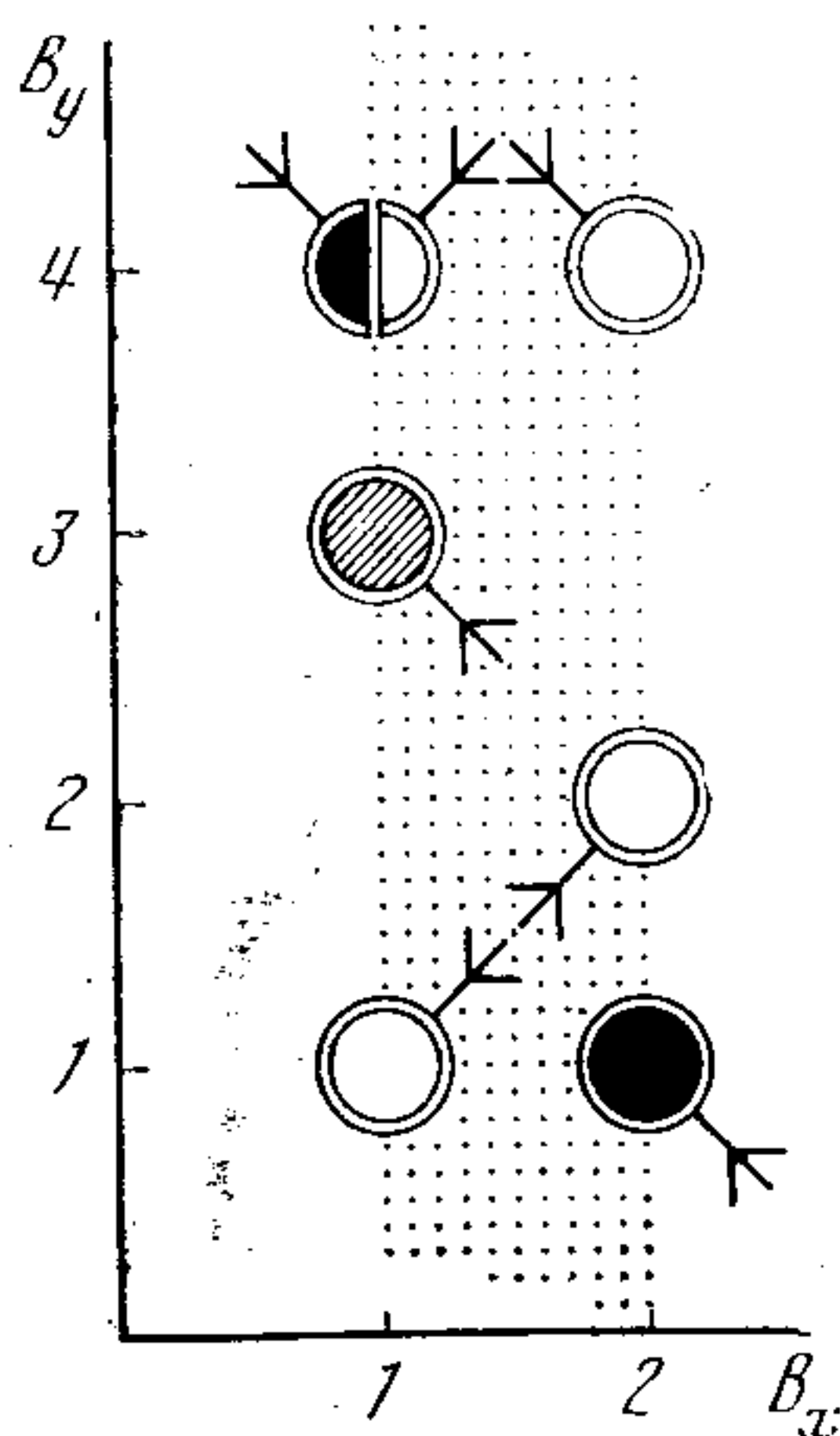


Рис. 3.
Пример 5

Обозначения те же, что и на рис. 2

Заметим, что попытка провести обобщение так же, как в предыдущем примере, ничего не даст, потому что применение правила 2 приводит к результату, для которого одно из противоречий является частным случаем. Это и говорит о том, что тесноты связок в выражениях не соответствуют искомой закономерности. Это очевидно и с содержательной точки зрения, так как закономерность состоит в соотношении между координатой положения системы по оси X и относительной координатой руки по той же оси, а в первоначальных записях они связаны самой рыхлой связью. Важно, однако, что сама система может определить несоответствие формы записи и отыскиваемой закономерности.

1. Итак, разумно применить правило 3 к каждому члену таблицы присутствия. Получаем

$$\Sigma_1 \equiv (B_x1 \wedge B_y1 \wedge P_x1 \wedge P_y1) \vee (B_x2 \wedge B_y2 \wedge P_x0 \wedge P_y0) \vee \\ \vee (B_x2 \wedge B_y4 \wedge P_x0 \wedge P_y1) \vee (B_x1 \wedge B_y4 \wedge P_x1 \wedge P_y1).$$

2. Теперь пробуем применить правило 2 к Σ_1 , отрывая набор $\{B_x\alpha_i\}$. Однако полученный результат

$$\Sigma_2 \equiv (B_x1 \vee B_x2 \vee B_x2 \vee 2 \vee B_x1) \wedge ((B_y1 \wedge P_x1 \wedge P_y1) \vee \\ \vee (B_y2 \wedge P_x0 \wedge P_y0) \vee (B_y4 \wedge P_x0 \wedge P_y1) \vee (B_y4 \wedge P_x1 \wedge P_y1))$$

приходится отвергнуть, так как $\tilde{S}_2 \rightarrow \Sigma_2$. Пробуем применить

правило 2 к Σ_1 , отрывая B_y . Получаем

$$\Sigma_3 \equiv (B_y1 \underset{3}{\vee} B_y2 \underset{3}{\vee} B_y4 \underset{3}{\vee} B_y4) \underset{2}{\wedge} ((B_x1 \underset{2}{\wedge} P_x1 \underset{2}{\wedge} P_y1) \underset{3}{\vee} \underset{3}{\vee} (B_x2 \underset{2}{\wedge} P_x0 \underset{2}{\wedge} P_y0) \underset{3}{\vee} (B_x2 \underset{2}{\wedge} P_x0 \underset{2}{\wedge} P_y1) \underset{3}{\vee} (B_x1 \underset{2}{\wedge} P_x1 \underset{2}{\wedge} P_y1)).$$

Этот результат ничему не противоречит.

3. Теперь к выражению $(B_y1 \underset{3}{\vee} B_y2 \underset{3}{\vee} B_y4 \underset{3}{\vee} B_y4)$ применяем правило 1 и получаем

$$\Sigma_4 \equiv (B_x1 \underset{2}{\wedge} P_x1 \underset{2}{\wedge} P_y1) \underset{3}{\vee} (B_x2 \underset{2}{\wedge} P_x0 \underset{2}{\wedge} P_y0) \underset{3}{\vee} (B_x2 \underset{2}{\wedge} \underset{2}{\wedge} P_x0 \underset{2}{\wedge} P_y1).$$

Этот результат тоже принимается.

4. Если мы теперь попробуем оторвать P_x , то получим такое Σ_5 , что $\tilde{S}_2 \rightarrow \Sigma_5$, значит этот путь закрыт. Но мы можем от Σ_4 оторвать P_y и получить

$$\Sigma_6 \equiv (P_y1 \underset{3}{\vee} P_y0 \underset{3}{\vee} P_y1) \underset{2}{\wedge} ((B_x1 \underset{2}{\wedge} P_x1) \underset{3}{\vee} (B_x2 \underset{2}{\wedge} P_x0) \underset{3}{\vee} \underset{3}{\vee} (B_x2 \underset{2}{\wedge} P_x0)).$$

5. Приводя подобные члены и заменяя $P_y0 \underset{3}{\vee} P_y1$ на 1, имеем окончательно обобщенную ситуацию

$$S \equiv (B_x1 \underset{2}{\wedge} P_x1) \underset{3}{\vee} (B_x2 \underset{2}{\wedge} P_x0).$$

Это означает, что положительный результат достигается при том условии, что система либо находится в точках с координатами B_x1 (левая граница полосы с травой), и ее рука направлена вправо (P_x1), причем вверх или вниз — не играет роли, либо находится в точках B_x2 (правая граница полосы), и рука направлена влево (P_x0). Это адекватное описание ситуации «травка под рукой».

Подчеркнем еще раз, что применение правила 3 вовсе не равносильно «оправданию» полного перебора гипотез о непривязанных наборах без учета структуры выражений. Дело в том, что

1) к правилу 3 мы прибегаем лишь в том случае, если уже установлена неадекватность записи и открываемого закона,

2) применение правила 3 не означает полного уничтожения структуры, а лишь «перемешивание» верхних слоев, сохраняющее все более глубокие связи.

Разумеется, разрушение структуры по причине отсутствия адекватности — путь, хотя иногда и необходимый, но, если его ничем не дополнить, нежелательный. Совершенно обязательно иметь методы, которые после открытия нового закона строили бы адекватную ему структуру. Для этого наш язык необходимо дополнить некоторыми новыми элементами. К сожалению, мы не

можем рассмотреть здесь связанные с этим вопросы. Большая их часть относится не к самому языку, а скорее, к его использованию системой.

В рамках системы, рассмотренной в [14], естественно считать, что задача обобщения поручена типовому блоку, обладающему полем внимания и долговременной памятью. Работа этого блока отнюдь не будет ограничиваться решением задачи обобщения начальных ситуаций, рассмотренной в этой статье, поскольку она является лишь частным случаем более сложных и важных задач обобщения в целом. Блок «обобщение» должен иметь специальный набор рецепторов, показания которых определяются не «внешним миром», как, скажем, показания рецептора «зрение», а зависят от содержимого поля внимания того блока, по заказу которого решается задача обобщения в данный момент. Алгоритмы обобщения будут закладываться в память блока в виде совокупности априорных фактов, определяющих выбор тех или иных операций — действий блока «обобщение». В частности, изложенные в настоящей статье правила обобщения можно рассматривать как некоторые из априорных фактов в памяти блока «обобщение».

Содержимое этой памяти может меняться в процессе работы блока, прежде всего в результате накопления экспериментальных фактов, содержащих описания примеров применения тех или иных операций. Кроме того, естественно потребовать, чтобы задача обобщения этих фактов могла решаться самим блоком «обобщение». Такая система, обладающая обучаемым обобщением, в процессе своей жизнедеятельности будет не только накапливать знания, но и улучшать способы их обработки.

Л и т е р а т у р а

1. М. М. Бонгард, И. С. Лосев, В. В. Максимов, М. С. Смирнов. Формальный язык описания ситуаций, использующий понятие связи.— Наст. сб.
2. Герц Г. Принципы механики, изложенные в новой связи. Изд-во АН СССР, М., 1959.
3. А. М. Касаткин. Моделирование сложного поведения с помощью М-автоматов.— Сб. «Механизмы и принципы целесообразного поведения», М., «Наука», 1972.
4. Н. Нильсон. Мобильный автомат, построенный с использованием принципов искусственного интеллекта.— Сб. «Интегральные роботы», под ред. Г. Е. Поздняка. М., «Мир», 1973.
5. И. М. Файгенберг. Мозг, психика, здоровье. М., «Наука», 1972.
6. Д. Юм. Исследование о человеческом познании. Соч., т. 2. М., «Мысль», 1965.
7. Ф. Бэкон. Сочинения, т. 2. М., «Мысль», 1972.
8. В. В. Максимов. Система, обучающаяся классификации геометрических изображений.— Наст. сб.
9. А. Эйнштейн, Л. Инфельд. Эволюция физики. ГТТИ, Л.— М., 1948.
10. И. М. Гельфанд, В. С. Гурфинкель, М. Л. Цетлин, М. Л. Шик. Некоторые вопросы исследования движений.— Сб. «Модели структурно-функциональной организации некоторых биологических систем». М., «Наука», 1966.

11. Логика и эмпирическое познание. Сб. под ред. П. В. Таванца. М., «Наука», 1972.
12. М. М. Бонгард, М. Н. Вайнцвайг, Ш. А. Губерман, М. Л. Извекова, М. С. Смирнов. Использование обучающейся программы для выявления нефтеносных пластов.— Геология и геофизика, 1966, № 6.
13. И. М. Гельфанд, Ш. А. Губерман, М. П. Житков, М. С. Калецкая, В. И. Кейлис-Борок, Е. Я. Ранцман, И. М. Ротвайн. Прогноз места возникновения сильных землетрясений как задача распознавания.— Наст. сб.
14. М. М. Бонгард, И. С. Лосев, М. С. Смирнов. Проект модели организации поведения — «Животное».— Наст. сб.

М. Н. Вайнцвайг, М. П. Полякова

ОБ ОДНОМ ПОДХОДЕ К ПРОБЛЕМЕ СОЗДАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

1. Что такое интеллект. Приступая к созданию систем искусственного интеллекта, необходимо прежде всего решить вопрос о том, какое из свойств естественного интеллекта следует считать определяющим. Чаще всего в качестве такого свойства рассматривают *способность решать широкий круг достаточно сложных задач*, поскольку сложностью и разнообразием решаемых задач обычно оценивается интеллектуальность человека. Эта точка зрения на интеллект согласуется и с практическими целями, а именно с желанием иметь устройства, способные выполнять функции специалистов в той или иной области. Распространение такой точки зрения привело к тому, что подавляющее большинство исследователей в области искусственного интеллекта пошло по пути построения систем, решающих некоторые классы достаточно сложных для человека задач, надеясь, по-видимому, что рано или поздно в принципах построения этих систем обнаружится нечто общее и это общее даст возможность построить модель интеллекта в целом. Так возникли программы для доказательства теорем, игры в шахматы, перевода с одного языка на другой, распознавания образов, а также различные системы типа «вопрос — ответ», «интеллектуальные» роботы и пр. Многие из этих работ быстро нашли практическое применение, что послужило стимулом для развития этого направления. Однако, если конечной целью таких исследований считать построение целостной модели интеллекта, то выбранный путь нам представляется неоправданно сложным и малоэффективным.

Дело прежде всего в том, что к моменту, когда человек способен решать сколько-нибудь сложные задачи, его память уже хранит чрезвычайно большое количество различных сведений, а