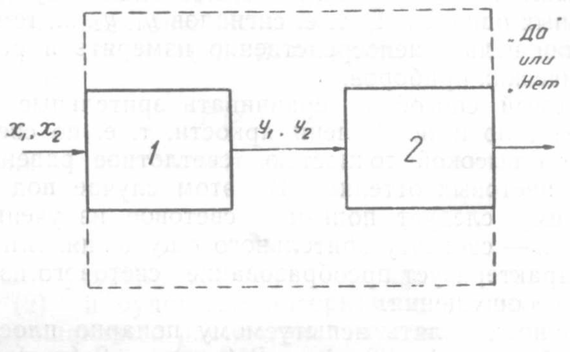
H9

О МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ УЗНАВАНИЯ

Ю. П. Шабанов-Кушнаренко, И. В. Шульгин, Б. К. Лопатченко

Известно, что человек может узнавать цвета и звуки, речь, рисунки, смысл фраз и многое другое. Очевидно, что узнавание представляет собой процесс преобразования информации, в основе которого лежат вполне определенные алгоритмы. Можно ли узнать что-либо достоверное об этих алгоритмах и каким образом?

В настоящей статье излагается один весьма общий подход к решению данной проблемы. Пусть имеется некоторый преобразователь информации *1*, алгоритм F которого необходимо расшифровать (см. рисунок). Сигналы, подаваемые на вход преобразователя, будем обозначать через $x_1, x_2, \ldots$. Возникающие при этом сигналы на выходе преобразователя обозначим через $y_1, y_2 \ldots$



Примем, что преобразователь *1* однозначный, однако не взаимно. Это означает, что выходные сигналы будут равны между собой не только в том случае, когда соответствующие им входные сигналы равны, но и в некоторых случаях, когда входные сигналы различны. В других же случаях различным выходным сигналам будут соответствовать разные выходные сигналы.

К выходу преобразователя *1* подключен преобразователь *2*, назначение которого состоит в следующем. Подадим на вход блока *1* поочередно сигналы x_1 и x_2. Под их действием на выходе этого блока возникнут сигналы y_1 и y_2. Блок *2* осуществляет сравнение сигналов y_1 и y_2 и в случае их равенства реагирует сигналом «да», в случае же неравенства — сигналом «нет».

Представим себе, что блоки *1* и *2* заключены в единый «черный ящик», обозначенный на рисунке пунктиром. Исследователь имеет возможность подавать на вход «черного ящика» пары различных входных сигналов x_1 и x_2 и, наблюдая реакцию системы в целом, судить о том, совпадают или нет при этом сигналы y_1 и y_2 на выходе блока *1*. Сами же сигналы y_1 и y_2 для непосредственного измерения недоступны.

Требуется по результатам эксперимента расшифровать вид алгоритма F, лежащего в основе работы блока *1*.

Замечательно, что описанный выше подход к проблеме узнавания распространяется на весьма широкий класс задач, имеющих прямое отношение к проблеме расшифровки алгоритмов узнавания. Мы полагаем, что такой подход дает возможность получить весьма обширную и притом вполне достоверную информацию об алгоритмах, реализуемых мозгом человека в процессе узнавания.

Рассмотрим примеры задач, сводящихся к этому подходу.

1. Известно, что различные световые излучения, воздействуя на орган зрения, вызывают в нашем сознании ощущение цвета. Одним и тем же излучениям всегда соответствуют одинаковые цвета. Разным излучениям, как правило, соответствуют различные цвета. Однако можно подыскать пары совершенно различных излучений, которые порождают одинаковые цвета. Испытуемый может очень точно установить, равны или не равны цвета, порождаемые парой тех или иных излучений.

В рассматриваемом случае под сигналами x_1, x_2,... можно понимать световые излучения; сигналам y_1, y_2,... соответствуют ощущения цвета, порождаемые этими излучениями в сознании испытуемого. Под реакцией системы в целом следует понимать ответ испытуемого, сообщающего о равенстве или неравенстве зрительных ощущений по цвету. Под алгоритмом F в данном случае понимается алгоритм преобразования светового излучения в цвет.

Важно отметить, что хотя мы не сомневаемся в существовании цветовых зрительных ощущений, т. е. сигналов y_1, y_2, ..., тем не менее, цвет этих ощущений нельзя непосредственно измерять и регистрировать с помощью физических приборов.

2. Испытуемый способен сравнивать зрительные ощущения не только по цвету, но и по степени яркости, т. е. по светлоте, отмечая со сравнительно высокой точностью светлотное равенство или неравенство двух цветовых оттенков. В этом случае под сигналами x_1, x_2,... по-прежнему следует понимать световое излучение, а под сигналами y_1, y_2, ... — светлоту зрительного ощущения. Алгоритм F в данном случае характеризует преобразование светового излучения в светлоту зрительного ощущения.

3. Будем предъявлять испытуемому попарно плоские ахроматические (черно-белые) фотографии $B_1(x, y)$ и $B_2(x, y)$, где $B(x, y)$ — — яркость зрительной картины в точке с координатами x, y. Предложим испытуемому реагировать ответом «да», если две предъявляемые фотографии по существу одинаковы, но отличаются лишь размером (масштабом), и ответом «нет», если фотографии отличаются чем-либо еще. Очевидно, что испытуемый весьма точно сможет произвести такую классификацию.

В данном случае под входными сигналами x_1, x_2 нужно понимать зрительные картины $B_1(x, y)$, $B_2(x, y)$, ..., а под выходными сигналами y_1, y_2, ... — представления испытуемого об этих картинах, абстрагированные от размера картин. Алгоритм F в рассматриваемом случае следует трактовать как алгоритм преобразования двумерной зрительной картины в представление определенного типа об этой картине.

4. Испытуемый способен также установить, одинаков ли смысл (содержание) двух различных слов, фраз или текстов. В этом случае мы приходим к задаче отыскания вида алгоритма F преобразования словесного материала в его смысловое содержание.

5. В различных жизненных ситуациях у человека возникают эмоции голода, страха, досады, удивления и т. д. Ощущая их, испытуемый легко устанавливает, одинаковые или различные эмоции возникали

у него в тех или иных ситуациях. Если бы удалось дать математическое описание объективных ситуаций, порождающих эмоции, мы получили бы возможность, придерживаясь принятого подхода, расшифровать алгоритм F преобразования жизненных ситуаций в эмоции.

Кроме перечисленных, можно сформулировать множество других задач, подобных приведенным выше, на которые распространяется предложенный нами подход.

Предлагаемая постановка проблемы, по-видимому, не охватывает всего комплекса задач, которые принято обозначать термином «узнавание». Объектом исследования здесь служат алгоритмы, объединяющие сигналы и классы. Такие алгоритмы, как известно, всегда участвуют в любом процессе узнавания.

Вырисовывается следующий ход исследования по расшифровке алгоритма F блока *1*. Выбираем некоторую конкретную задачу для решения ее испытуемым и первоначально накапливаем определенный экспериментальный материал. Для этого испытуемому предъявляем различные пары входных сигналов x_1, x_2 и протоколируем ответ испытуемого («да» или «нет»). В один класс относим все пары входных сигналов, которые порождают ответ «да», а в другой — все пары порождающие ответ «нет». Далее на базе этого эмпирического материала формулируем некоторое высказывание $A(x_1, x_2)$ истинное для всех пар входных сигналов первого класса (ответ «да») и ложное для всех пар второго класса (ответ «нет»).

В результате получаем экспериментальный закон следующего вида: «Если и только если $A(x_1, x_2)$ истинно, то выходные сигналы y_1, y_2 блока *1* совпадают, т. е.

$$F(x_1) = F(x_2)''. \tag{1}$$

Дальнейшая задача состоит в том, чтобы отыскать общий вид оператора

$$y = F(x), \tag{2}$$

удовлетворяющий условию 1.

Формула (2) и будет тем алгоритмом, который лежит в основе работы мозга, решающего рассматриваемую задачу. Поскольку утверждения (1) и (2) равносильны, формула (2) содержит всю ту информацию, которую можно получить, действуя в пределах рассматриваемой постановки проблемы. Не выходя за эти пределы, получить какую-либо дополнительную информацию о работе мозга невозможно.

Сформулируем следующую теорему: *«Если $G(x)$ является частным видом оператора, удовлетворяющим условию (1), то общий вид оператора $F(x)$, удовлетворяющий тому же условию, запишется в форме суперпозиции оператора G и произвольной взаимно-однозначной зависимости f, т. е.*

$$F(x) = f[G(x)]''. \tag{3}$$

Значение этой теоремы состоит в следующем. Предположим, что нам удалось найти конкретный алгоритм $G(x)$, удовлетворяющий условию (1), т. е. такой, что для него справедливо утверждение: «Если и только если $A(x_1, x_2)$ истинно, то

$$G(x_1) = G(x_2)''. \tag{4}$$

Возникает вопрос, можно ли считать оператор $G(x)$ именно тем, алгоритмом, который в действительности реализован мозгом. Этого с уверенностью нельзя сказать, так как не исключено, что найдется другой алгоритм, отличный от алгоритма $G(x)$ и также удовлетворяющий условию (2).

Относительно же оператора $f[G(x)]$ можно с уверенностью сказать, что он включает и тот конкретный алгоритм, который реализован мозгом, так как теоремой доказывается, что других операторов удовлетворяющих условию (2), не существует. Это значит, что всегда можно так подобрать зависимость f, чтобы оператор $f[G(x)]$ совпадал с алгоритмом, реализуемым человеческим мозгом.

Вместе с тем следует особо подчеркнуть тот факт, что в пределах принятой нами постановки проблемы конкретный вид зависимости f не может быть расшифрован. Для этого необходимы исследования, проводимые по принципиально иной методике.

Таким образом, сформулированная теорема позволяет свести задачу отыскания общего вида оператора, удовлетворяющего условию (1), к более простой задаче нахождения хотя бы одного частного вида оператора, удовлетворяющего тому же условию.

Приступим к доказательству теоремы.

Пусть $G(x)$ есть частный вид оператора, удовлетворяющий условию (4). Предположим, что существует другой оператор $F(x)$, удовлетворяющий условию (1). Выберем сигналы x_1 и x_2 так, чтобы $G(x_1) = G(x_2)$. Согласно условию (4) высказывание A (x_1, x_2) истинно. Следовательно, в силу условия (1) $F(x_1) = F(x_2)$. Выберем теперь сигналы x_1 и x_2 так, чтобы выполнялось условие $F(x_1) = F(x_2)$. Согласно (1) высказывание $A(x_1, x_2)$ истинно, следовательно, в силу утверждения (4) $G(x_1) = G(x_2)$.

Отсюда следует, что сигналы $F(x)$ и $G(x)$ должны быть связаны некоторой взаимно однозначной зависимостью f.

Таким образом, любой оператор $F(x)$, удовлетворяющий условию (1), записывается в виде выражения (3).

Докажем теперь, что в качестве f может быть принята любая взаимно однозначная зависимость. Пусть f — произвольная взаимно однозначная зависимость, а $G(x)$ — оператор, удовлетворяющий условию (4). Тогда оператор $f[G(x)]$ также должен удовлетворять условию (1), т. е. будет справедливо следующее утверждение: «Если и только если $A(x_1, x_2)$ истинно, то

$$f[G(x_1)] = f[G(x_2)]. \tag{5}$$

Действительно, пусть $A(x_1, x_2)$ истинно. Тогда, согласно (4), $G(x_1) = G(x_2)$. Следовательно, $f[G(x_1)] = f[G(x_2)]$. Пусть теперь $f[G(x_1)] = f[G(x_2)]$. Это значит, что $G(x_1) = G(x_2)$. Следовательно, в соответствии с (4), $A(x_1, x_2)$ истинно, и справедливость утверждения (5) доказана.

Итак, оператор $F(x) = f[G(x)]$ действительно удовлетворяет условию (1). Вместе с тем других операторов, удовлетворяющих условию (1), не существует. Теорема доказана.