ВОПРОСЫ ОБРАБОТКИ ИНФОРМАЦИИ ПРИ ПРОЕКТИРОВАНИИ СИСТЕМ
(Вычислительные системы)
1977 год Выпуск 69

УДК 51.16

ОБ ОДНОЙ ИНДУКТИВНОЙ ЗАДАЧЕ ПРЕДСКАЗАНИЯ

Б.П.Гавриленко

Известно, что в распознавании образов приходится иметь дело с различными типами задач (см., например, [1]), для каждого из которых существуют свои методы решения. Среди этих типов задач, пожалуй, наиболее важной является задача поиска решающих функций. Зачастую оказывается, что мы вынуждены решать задачи некоторых других типов только затем, чтобы более успешно решить ту или иную задачу поиска решающей функции.

Знание решающей функции позволяет по известным значениям признаков произвольного контрольного объекта предсказать, к ка - кому из образов он относится. Можно говорить, что такая информация о контрольном объекте является целевой в задачах этого типа. Следует отметить, что во всякой такой задаче распознавания образов обучающая выборка содержит сведения о том, к какому из обра - зов относится каждый ее объект. Эти сведения получают с помощью тех или иных измерений, результаты которых всегда представлены в шкале наименований (формулировки понятия шкалы и других понятий теории измерений содержатся в [2]). Таким образом, во всякой задаче распознавания рассматриваемого типа целевым признаком является признак, измеряемый в шкале наименований.

Из этих рассуждений следует, что задача распознавания такого типа является частным случаем, например, такой более общей задачи, когда целевой признак может быть представлен любой шкалой. (Эта особенность задачи распознавания отмечается в работе [3, с.68-69].) Кроме того, следует иметь ввиду, что не всегда результат измерения какого-либо свойства некоторого объекта может быть представ - лен в виде числа или символа. Так, в задачах ситуационного управ-

ления приходится иметь дело с такими результатами наблюдения или, что то же самое, измерения свойств объектов, когда эти результаты представлены, например, графами или отношениями на элементах измеряемого объекта. Для такого рода измерений не представляется возможным предложить какую-либо шкалу в смысле работы [2]. Но если шкалу интерпретировать как средство, предоставляющее возможность формально выражать некоторую "семантику" результатов измерения, то это позволяет с единой точки зрения рассматривать более широкий, чем в [2], класс измерений. (Один из аспектов такого подхода к измерениям будет затронут в настоящей работе.) Задачи на предсказание значения целевого признака контрольного объекта, в которых как целевому, так и нецелевым признакам могут соответствовать измерения из упомянутого более широкого класса, будут именоваться здесь задачами предсказания. В данной работе рассматривается только одна из возможных задач предсказания.

Необходимо также отметить, что в известных из литературы постановках задач распознавания не задается информация о "семантике" (например, о шкалах) используемых признаков. В результате, как правило, наблюдаются следующие две ситуации.

I. Предлагаются алгоритмы, которые заранее предназначены для решения фиксированного класса задач. Для этих задач обучающие выборки и контрольные объекты необходимо задавать только в таких признаковых пространствах, тип которых фиксирован для каждого конкретного алгоритма. Такая ситуация не всегда осознается авторами предлагаемых алгоритмов, а тем более теми, кто применяет их для решения практических задач.

2. Во многих алгоритмах предусматривается использование некоторой меры сходства между объектами. Но выбор той или иной меры сходства для решения конкретной задачи зачастую не оговаривается и зависит от опыта и интуиции исследователя.

В настоящей работе предложен алгоритм решения задачи предсказания, для которого информацию о содержательных свойствах используемых признаков необходимо сообщать явно, в виде алгоритмов извлечения "семантической" информации из эмпирического материала.

§I. Исходные понятия и обозначения

Введем некоторые формальные понятия и обозначения, необходимые для точной формулировки индуктивной задачи предсказания.

Мы будем иметь дело с объектами, которые обнаруживаются с помощью некоторого набора измерительных процедур. С каждой измерительной процедурой из этого набора будет связано понятие признака. Таким образом, каждая измерительная процедура предназначается для определения некоторого значения соответствующего признака на измеряемом объекте.

Символом x_j^i будем обозначать значение i-го признака на j-м объекте. В качестве значений признаков могут выступать символы, числа, последовательности символов, кодирующие наблюдаемые отношения на элементах объекта в том случае, когда объект состоит из элементов.

Если n равно количеству используемых признаков, то упорядоченная последовательность вида $\langle x_j^1, \ldots, x_j^n \rangle$ будет задавать j-й объект. Каждую такую последовательность будем обозначать символом O с некоторым индексом. Конечную совокупность $\{O_j\}$ некоторых объектов назовем обучающей выборкой.

Контрольным объектом будем называть упорядоченную последовательность вида $\langle x_j^1, \ldots, x_j^{i-1}, x_j^{i+1}, \ldots, x_j^n \rangle$, т.е. такую, в которой отсутствует значение i-го признака. Признак, значение которого отсутствует в контрольном объекте, будем называть целевым.

С каждым признаком мы будем связывать определенный алгоритм, формально выражающий некоторые содержательные свойства этого признака. Для того чтобы пояснить, какие свойства здесь имеются в виду, воспользуемся следующим примером. Пусть в результате некоторых измерений определенного признака на четырех объектах были получены в качестве значений этого признака числа 2, 4, 2 и 5 соответственно. Очевидно, что смысл этих символов будет различным в зависимости от того, в какой шкале выражены эти результаты. Если эти числа получены в результате выяснения профессии каких-то четырех человек, то наша интуиция легко отличит их смысл от смысла тех же символов, но полученных для этих же четырех человек в результате "измерения" их успеваемости по некоторому предмету. В первом случае эти символы говорят нам только о том, что профессии совпадают лишь у первого и третьего человека. Во втором же случае нам известно не только то, что успеваемость у первого и третьего из них совпадает, но также и то, что оба они освоили предмет хуже второго и все трое освоили этот предмет хуже четвертого.

Для того чтобы формально выражать смысл, содержащийся в произвольной конечной совокупности значений произвольного признака, мы введем понятие алгоритма извлечения "семантической" информации или, что проще, семантического алгоритма. Таким образом, чтобы формально задать содержательные свойства некоторого признака, необходимо указать соответствующий этому признаку семантический алгоритм.

Для всякой упорядоченной последовательности значений $x^1_{j_1},\ldots, x^1_{j_n}$ признака 1 через $px^1_{j_1},\ldots, px^1_{j_n}$ будет обозначаться последовательность, полученная из исходной некоторой перестановкой p ее элементов.

С е м а н т и ч е с к и м а л г о р и т м о м для признака 1 будем называть алгоритм f_1, обладающий по крайней мере следующими свойствами:

а) алгоритм f_1 применим ко всякой непустой конечной упорядоченной совокупности значений признака 1, и на каждой такой совокупности его значением является некоторый код, не обязательно различный для отличающихся совокупностей;

б) для всяких значений a и b алгоритма f_1, для всяких упорядоченных совокупностей $x^1_{i_1},\ldots, x^1_{i_n}$ и $x^1_{j_1},\ldots, x^1_{j_m}$ значений признака 1, если $f_1(x^1_{i_1},\ldots, x^1_{i_n}) = a$, $f_1(x^1_{j_1},\ldots, x^1_{j_m}) = b$ и $n \neq m$, то $a \neq b$;

в) для произвольной перестановки p, для произвольных последовательностей $x^1_{i_1},\ldots, x^1_{i_n}$, $x^1_{j_1},\ldots, x^1_{j_n}$ и $x^1_{k_1},\ldots, x^1_{k_n}$ значений признака 1 таких, что $f_1(x^1_{j_1},\ldots, x^1_{j_n}) = f_1(x^1_{k_1},\ldots, x^1_{k_n})$ и $f_1(x^1_{i_1},\ldots, x^1_{i_n}) = f_1(px^1_{j_1},\ldots, px^1_{j_n})$, должно выполняться $f_1(x^1_{i_1},\ldots, x^1_{i_n}) = f_1(px^1_{k_1},\ldots, px^1_{k_n})$.

Приведем примеры возможных семантических алгоритмов для некоторых (широко распространенных в распознавании образов) типов признаков, свойства каждого из которых характеризуются определенным типом шкалы.

Пусть одномерный массив a, состоящий из n ячеек, задает совокупность значений некоторого признака. Тогда такой признак может характеризоваться следующими алгоритмами, написанными на языке АЛГОЛ в виде процедур-функций.

I. Шкала наименований. Процедурой восстанавливается отношение равенства на множестве значений признака, которое задается массивом a. Поскольку совокупность значений признака упорядоченна, то матрица размерности n x n истинностных значений выражений вида $x_i = x_j$, которая могла бы быть при этом получена, кодируется одним числом, способ получения которого довольно тривиален.

```
real procedure f1(a,n); array a; integer n;
begin integer i,j,k; real b; b:= k:= 0;
for i:= 1 step 1 until n-1 do
for j:= i+1 step 1 until n do begin
if a[i]= a[j] then b:= b+2 ↑ k;
k:= k+1 end; b:= b+2↑k ;
f1:= b end процедуры;
```

II. Шкала порядка. Процедурой восстанавливается отношение порядка на множестве значений признака.

```
real procedure f2(a,n); array a; integer n;
begin integer i,j,k; real b; b:= k:= 0;
for i:= 1 step 1 until n do
for j:= 1 step 1 until n do begin
if i≠j then begin
if a[i] ≤ a[j] then b:= b+2↑k ;
k:= k+1; end end; b:= b+2↑k ;
f2:= b end процедуры;
```

III. Шкала интервалов. Нижеследующий алгоритм, как можно заметить, только частично передает свойства шкалы интервалов, так как в нем реализовано восстановление всего лишь одного двухместного отношения $x \leq y$ и одного трехместного отношения $\frac{x+y}{2} \leq z$. Очевидно, что эту шкалу можно характеризовать алгоритмом, который восстанавливает также и другие отношения, в частности, большей местности. Тем не менее он вполне пригоден для практического использования.

```
real procedure f3(a,n);
array a; integer n;
begin integer i,j,k,l; real b; b:= l:= 0;
```

```
for i:= 1 step 1 until n do
for j:= 1 step 1 until n do
if i≠j then begin
if a[i]≤ a[j] then b:= b+2↑1 ;
1:= 1+1 end;
for k:= j+1 step 1 until n do begin
if (a[k] + a[j])/2 ≤ a[i] then b:= b+2↑1 ;
1:= 1+1 end end;
b:= b+2↑1 ; f3:= b end процедуры;
```

IУ. Шкала отношений. В приводимом алгоритме восстанавливается отношение порядка и два трехместных отношения: $\frac{x+y}{2} \leq z$ и $x+y \leq z$. Таким образом, свойства шкалы отношений алгоритм передает только частично.

```
real procedure f4(a,n); array a; integer n;
begin integer i,j,k,1; real b; b:= 1:= 0;
for i:= 1 step 1 until n do
for j:= 1 step 1 until n do begin
if i≠j then begin
if a[i]≤ a[j] then b:= b+2↑1 ;
1:= 1+1 end;
for k:= j+1 step 1 until n do begin
if (a[k] + a[j])/2 ≤ a[i] then b:= b+2↑(1+1);
if a[k] + a[j] ≤ a[i] then b:= b+2↑1 ;
1:= 1+2 end end;
b:= b+2↑1 ; f4:= b end процедуры;
```

Доказательство того, что приведенные алгоритмы удовлетворяют условиям "а"-"в" определения семантического алгоритма, мы опускаем ввиду его простоты.

Помимо семантического алгоритма, для целевого признака будет задаваться также некоторый дополнительный алгоритм - "генератор гипотез". При определении этого алгоритма символом X^1 мы будем обозначать множество тех и только тех значений признака 1 ,которые встречаются в обучающей выборке $\{O_j\}$.

Г е н е р а т о р о м г и п о т е з для (целевого) признака 1 будем называть алгоритм Γ_1 , удовлетворяющий следующим условиям:

а) алгоритм Γ_1 определен на произвольной упорядоченной паре $\langle X^1, f_1 \rangle$, и на всякой такой паре его значением является ко -

нечное непустое множество $\{G^1_k\}$, каждый элемент G^1_k которого представляет собой непустое множество значений признака 1, т.е. $\Gamma_1(\langle X^1, f_1\rangle) = \{G^1_k\}$;

б) пересечение произвольных двух элементов G^1_i и G^1_j множества $\{G^1_k\}$ является пустым;

в) если произвольные последовательности $x^1_{i_1},\ldots, x^1_{i_n}$ и $x^1_{j_1},\ldots, x^1_{j_n}$ значений признака 1 таковы, что $x^1_{i_1}, x^1_{j_1} \in G^1_s, \ldots, x^1_{i_n}, x^1_{j_n} \in G^1_t$, причем $G^1_s,\ldots, G^1_t$ являются элементами множества $\{G^1_k\}$, то для всякого a $f_1(x^1_{i_1},\ldots, x^1_{i_n}) = a$ тогда и только тогда, когда $f_1(x^1_{j_1},\ldots, x^1_{j_n}) = a$.

Если для уже упоминавшихся типов шкал в качестве семантических алгоритмов фиксировать приведенные нами процедуры f1,f2,f3 и f4 , то генераторы гипотез для признаков, измеряемых в этих шкалах, будут состоять в следующем.

I. Шкала наименований. Для нее множество $\{G^1_k\}$ будет состоять из всех возможных значений соответствующего признака, т.е. каждое множество G^1_k , единственным элементом которого является некоторое значение признака 1 , будет являться элементом множества $\{G^1_k\}$.

II. Шкала порядка. Пусть для измеряемого в шкале порядка признака 1 множество X^1 его значений, встречающихся в обучающей выборке, состоит из элементов $x^1_1, x^1_2, \ldots, x^1_n$. Тогда множество $\{G^1_k\}$ будет состоять из следующих элементов. Первый элемент будут составлять все те возможные значения x^1_i признака, для которых выполнено условие $x^1_i < x^1_1$. Второй элемент будет состоять из единственного значения x^1_1 . В третий элемент войдут все те значения x^1_i признака, которые удовлетворяют условию $x^1_1 < x^1_i < x^1_2$. Четвертый элемент - это значение x^1_2 и т.д. Предпоследний элемент будет состоять из единственного значения x^1_n . Последний (2n+1)-й элемент множества $\{G^1_k\}$ - это все те значения x^1_i признака, для которых выполнено условие $x^1_n < x^1_i$.

Ш. Шкала интервалов. Множество X^1 значений признака, встречающихся в обучающей выборке, пополняется новыми элементами, которые вычисляются по формулам $x = 2x_i - x_j$ и $x = (x_i + x_j)/2$, где x_i и x_j - значения признака, в роли которых выступают поочередно все элементы множества X^1 . Из всех попарно неодинаковых элемен-

тов полученного таким образом множества формируется вспомогательное множество X^*. Построение множества $\{G_k^i\}$, отправляясь от множества X^*, осуществляем, как для шкалы порядка.

IУ. Шкала отношений. Отличие от предыдущего примера состоит лишь в том, что множество X^i значений признака, встречающихся в обучающей выборке, пополняется еще и элементами, вычисляемыми по формуле $x = x_i/2$, где в качестве x_i берутся поочередно все элементы множества X^i, а также по формуле $x = x_i + x_j$, где значения символов x_i и x_j совпадают со значениями этих же символов для формул предыдущего примера.

Еще раз отметим, что как семантические алгоритмы, так и генераторы гипотез фиксируют наши предположительные знания (т.е. гипотезы) о свойствах соответствующих признаков.

§2. Постановка задачи

Использовав введенные понятия, сформулируем индуктивную задачу предсказания следующим образом.

Фиксируется совокупность некоторых признаков, количество которых равно n. Относительно каждого признака предполагается известным его семантический алгоритм f_s $(s = 1, \ldots, n)$.

Имеется обучающая выборка $\{O_j\}$, состоящая из m объектов, каждый из которых задан последовательностью значений указанных признаков.

Задан контрольный объект K, для которого неизвестно значение i-го признака, являющегося целевым $(1 \le i \le n)$.

Для целевого признака фиксирован также генератор гипотез Γ_i.

Упорядоченную последовательность $\langle \{O_j\}, K, i, \{f_s\}, \Gamma_i \rangle$ будем называть входной информацией и обозначать символом $\mathcal{I}$.

Для контрольного объекта K требуется указать, использовав $\mathcal{I}$, такое множество g значений целевого признака i, чтобы

$$\forall x (x \in g \to [\exists G_j^i (x \in G_j^i \,\&\, G_j^i \in \{G_k^i\})]),$$

где $\{G_j^i\} = \Gamma_i(\langle X^i, f_i \rangle)$, т.е. множество, являющееся подмножеством множества значений целевого признака, порождаемого генератором гипотез для этого признака. В этом состоит рассматриваемая здесь индуктивная задача предсказания.

В дальнейшем нас будут интересовать не все возможные алгоритмы ее решения, т.е. алгоритмы вида $A_t(\mathcal{I}) = g$, а только такие, ко-

торые обладают некоторыми желательными свойствами. В работе [4] были сформулированы требования к алгоритмам решения несколько иной задачи предсказания. Тем не менее аналогичные требования следует предъявить, как мы считаем, и к интересующим нас алгоритмам. Для этого необходима точная формулировка этих требований, соответствующая рассматриваемой задаче предсказания.

Требование универсальности.

Алгоритм A_t должен быть применим к любой входной информации рассматриваемого здесь вида, т.е.

$$\forall \mathcal{I}_s \; A_t(\mathcal{I}_s) = g_{ts}. \qquad (2.1)$$

Выдвигая это требование, мы желаем иметь дело только с такими алгоритмами, возможность применения которых к той или иной задаче рассматриваемого вида не зависела бы от конкретной задачи предсказания, например, в каких шкалах измерены значения используемых в этой задаче признаков и т.п.

Требование нетривиальности.

Для всякого A_t должна существовать такая входная информация $\mathcal{I}_s = \langle \{O_j\}, K, i, \{f_s\}, \Gamma_i \rangle$, чтобы

$$g_{ts} \subset \{G_k^i\}, \qquad (2.2)$$

где $g_{ts} = A_t(\mathcal{I}_s)$, а $\{G_k^i\} = \Gamma_i(\langle X^i, f_i \rangle)$.

Требование нетривиальности ограничивает класс возможных алгоритмов только такими, каждый из которых (по крайней мере для некоторой соответствующей ему входной информации) позволяет указать такое множество значений целевого признака, которое является собственным подмножеством множества значений, определяемого генератором гипотез. Отметим, что относительно каждого значения целевого признака, принадлежащего какому-либо элементу множества $\{G_k^i\}$ и не являющегося элементом множества g_{ts}, утверждается: ни одно из них не появится фактически, если измерить целевой признак для контрольного объекта. Вместе с тем все эти значения генератором гипотез указываются как возможные. Очевидно, что получить дедукцией такое множество "запрещенных" значений из сведений, содержащихся во входной информации, невозможно. Здесь принципиально необходим тот или иной индуктивный шаг.

Требование содержательной инвариантности. Прежде чем сформулировать третье требование

к алгоритмам следует заметить, что входная информация $\mathcal{J}$ для интересующих нас алгоритмов содержит результаты некоторых наблюдений (т.е. измерений), выражаемые обучающей выборкой $\{O_j\}$ и контрольным объектом K. Кроме того, она содержит наши предположения (гипотезы) о свойствах используемых признаков, а именно алгоритмы $\{f_i\}$ и Γ.

Вообще говоря, все эти сведения можно было бы зафиксировать в виде некоторой другой входной информации, скажем $\mathcal{J}^*$, используя для записи результатов измерений и указанных предположений, допустим, другие символы. При этом $\mathcal{J}^*$ не отличалась бы от $\mathcal{J}$ с содержательной точки зрения. В связи с этим мы желаем потребовать от алгоритмов решения задачи предсказания, чтобы они обладали свойством давать "содержательно" одинаковые результаты для "содержательно" одинаковых входных данных.

Для того чтобы точно сформулировать требование содержательной инвариантности, введем некоторые понятия.

Предположим, что каждое значение x_j^i i-го признака является элементом некоторого множества символов W^i (быть может, бесконечного). Относительно ситуации, когда в качестве значений признаков выступают записи отношений на элементах измеряемого объекта, отметим следующее. Обобщение дальнейшего изложения и на этот случай не является сложным и не внесло бы ничего нового в рассматриваемую задачу. Но поскольку такое обобщение является технически более громоздким, то отмеченный случай рассматриваться не будет.

Для совокупности множеств $\{W^i\}$ $(i = 1,\dots, n)$ зафиксируем некоторую другую совокупность множеств символов, скажем $\{\hat{W}^k\}$ $(k = 1,\dots, n)$, такую, что каждому множеству W^i из первой совокупности можно взаимно-однозначно сопоставить некоторое множество $\hat{W}^k$ из второй совокупности, причем мощности множеств W^i и $\hat{W}^k$ совпадают. Предположим также, что существует совокупность $\{\alpha^i\}$ $(i = 1,\dots, n)$ произвольных эффективных взаимно-однозначных отображений таких, что каждое α^i всякому элементу множества W^i ставит в соответствие элемент множества $\hat{W}^k$.

Наличие $\{W^i\}$, $\{\hat{W}^k\}$ и $\{\alpha^i\}$ позволяет для описания результатов измерений в виде обучающей выборки и контрольного объекта использовать символы либо из множеств $\{W^i\}$, либо из $\{\hat{W}^k\}$. Более того, выбор той или иной совокупности множеств для этой цели никак не предпочтителен с эмпирической точки зрения.

Так как входная информация содержит гипотезы о свойствах используемых признаков, то соответствующие сведения должны формулироваться как при использовании $\{W^i\}$, так и при использовании $\{\hat{W}^k\}$, а также быть взаимно транслируемыми. Поэтому необходимо рассмотреть преобразования алгоритмов $\{f_i\}$ и Γ_i.

Будем говорить, что преобразование β^i $(i=1,\dots, n)$ является **допустимым для алгоритма** f_i, если оно представляет собой вычислимую функцию от алгоритма f_i, определенного на совокупностях значений признака из W_i, ставит ему в соответствие некоторый семантический алгоритм $\hat{f}_k$ (т.е. $\beta^i f_i = \hat{f}_k$), определенный на совокупностях значений признака из $\hat{W}^k$ и такой, что при этом выполняется условие: для всяких $x_1 \in W^i, \dots, x_\xi \in W^i$, $y_1 \in W^i, \dots, y_\xi \in W^i$ должно выполняться $f_i(x_1,\dots,x_\xi) = f_i(y_1,\dots,y_\xi)$ тогда и только тогда, когда

$$\hat{f}_k(\alpha^i x_1,\dots,\alpha^i x_\xi) = \hat{f}_k(\alpha^i y_1,\dots,\alpha^i y_\xi).$$

Условимся символами $\{\hat{O}_j\}$ и $\hat{K}$ обозначать такие обучающую выборку и контрольный объект соответственно, которые образуются из $\{O_j\}$ и K при переходе от $\{W^i\}$ к $\{\hat{W}^k\}$ с помощью совокупности отображений $\{\alpha^i\}$. При этом $\alpha^i(Z^i)$, где Z^i – некоторое множество значений какого-либо признака (т.е. подмножество одного из множеств W^i), будет обозначать такое подмножество $\hat{Z}^k$ соответствующего множества $\hat{W}^k$, которое образуется из множества Z^i применением преобразования α^i; множество, каждый k-й элемент которого представляет собой множество $\alpha^i(Y_k)$, будем обозначать через $\alpha^i[\{Y_k\}]$.

Преобразование γ^i является **допустимым для алгоритма** Γ_i, если оно представляет собой вычислимую функцию от алгоритма Γ_i, определенного на упорядоченных последовательностях вида $\langle X^i, f_i\rangle$, ставит ему в соответствие некоторый алгоритм $\hat{\Gamma}_k$ (т.е. $\gamma^i\Gamma_i = \hat{\Gamma}_k$), определенный на упорядоченных последовательностях вида $\langle \alpha^i(X^i), \beta^i f_i\rangle$ (т.е. $\langle \hat{X}^k, \hat{f}_k\rangle$) и такой, что при этом выполняется условие: для всяких X^i, f_i и $\{G_i^i\}$ должно выполняться $\Gamma_i(\langle X^i, f_i\rangle) = \{G_i^i\}$ тогда и только тогда, когда

$$\gamma^i\Gamma_i(\langle \alpha^i(X^i), \beta^i f_i\rangle) = \alpha^i[\{G_i^i\}].$$

Преобразования $\{\alpha^i\}$, $\{\beta^i\}$ и γ^i позволяют взаимно-однозначно перейти от входной информации $\mathcal{J} = \langle\{O_j\}, K, i, \{f_i\}, \Gamma_i\rangle$,

полученной с использованием $\{W^i\}$, к входной информации

$$\hat{\mathcal{J}} = \langle\langle \hat{O}_j \rangle, \hat{K}, n, \langle \hat{f}_x \rangle, \hat{\Gamma}_n \rangle ,$$

полученной с использованием $\{\hat{W}^k\}$. Здесь $\hat{f}_x = \beta^i f_i$ и $\hat{\Gamma}_n = \gamma^i \Gamma_i$.

Очевидно, что всякий раз, когда один исследователь, использующий $\{W^i\}$, запишет некоторые результаты измерений, определенные гипотезы о свойствах используемых признаков и цель в виде входной информации $\mathcal{J}$, то другой исследователь, использующий $\{\hat{W}^k\}$, при этом будет записывать те же результаты измерений, гипотезы и цель в виде входной информации $\hat{\mathcal{J}}$. Поэтому кажется разумным следующее

Требование содержательной инвариантности:

Для всякого алгоритма A_t, для всякой входной информации $\mathcal{J}$ и для всякого множества ε_{ts} должно выполняться $A_t(\mathcal{J}_s) = \varepsilon_{ts}$ тогда и только тогда, когда $A_t(\hat{\mathcal{J}}_s) = \hat{\varepsilon}_{ts}$, где $\hat{\varepsilon}_{ts} = \alpha^i[\varepsilon_{ts}]$. (2.3)

Это требование, как может показаться, можно было бы усилить. Для этого следовало бы позволить изменять входную информацию не только преобразованиями $\{\alpha^i\}$, $\{\beta^i\}$ и γ^i, но также и такими, с помощью которых можно было бы "склеивать" несколько признаков в один новый, а семантическим алгоритмам и генераторам гипотез "склеиваемых" признаков ставить в соответствие один семантический алгоритм и генератор гипотез для нового признака. Легко заметить, что в таких преобразованиях не всегда может участвовать целевой признак. В самом деле, поскольку для контрольного объекта значение целевого признака неизвестно, то для этого же объекта будут неизвестными и все те новые признаки, в образовании которых участвует целевой признак. Тем не менее преобразования, которые всякому старому значению целевого признака ставят во взаимно-однозначное соответствие несколько значений новых целевых признаков, а нецелевые признаки при этом могут заменяться такой совокупностью новых нецелевых признаков, что их количество может не совпадать с числом старых нецелевых признаков, рассматривать можно. Более того, ниже будет предложен алгоритм решения индуктивной задачи предсказания, содержательно инвариантный и относительно таких преобразований. Но все преобразования этого типа недопустимы в том случае, когда нужно не только предсказать значение целевого признака для контрольного объекта, но также использовать

для такого предсказания как можно меньшее количество признаков. При этом информация о том, что какой-либо признак оказался "неинформативным", является результирующей в задачах такого рода. Если ее получает один исследователь (пользующийся одними формальными средствами), то содержательно точно такую же информацию должен получать и всякий другой исследователь (использующий некоторые другие формальные средства), решающий содержательно ту же задачу. Именно поэтому преобразования, "склеивающие" признаки, следует считать недопустимыми.

§3. Алгоритм решения задачи

Рассмотрим пример алгоритма решения индуктивной задачи предсказания, удовлетворяющего приведенным требованиям.

Пусть фиксирована входная информация

$$\mathcal{J}_s = \langle\langle O_j \rangle, K, i, \langle f_i \rangle, \Gamma_i \rangle ,$$

где $j = 1, \ldots, m$; $1 = 1, \ldots, n$; $1 \le i \le n$.

Алгоритм A_0 состоит в следующем.

ШАГ 1. Генератором гипотез Γ_i порождаются все возможные множества $\{G^i_k\} = \Gamma_i(\langle X^i, f_i \rangle)$ значений целевого признака для контрольного объекта K. Из каждого множества выбирается произвольным образом по одному элементу. Все выбранные элементы образуют конечное множество, которое будет обозначаться символом h.

ШАГ 2. Рассматриваются поочередно все элементы множества h. Для каждого элемента обучающая выборка $\{O_j\}$ дополняется таким объектом, который образуется из контрольного объекта заполнением неизвестного значения целевого признака рассматриваемым элементом множества h. При этом для каждого элемента множества h образуется некоторое множество $\mathcal{H}_k$ объектов. Мощность каждого из множеств $\mathcal{H}_k$ равна $n+1$.

ШАГ 3. Для каждого из множеств $\mathcal{H}_k$ строятся все возможные упорядоченные последовательности объектов, имеющие длину N (число N равно числу обращений к шагу 3, увеличенному на единицу, так что, например, при первом обращении $N = 2$, при втором – $N = 3$ и т.д.).

Каждой ξ-й последовательности объектов сопоставляется строка, получаемая следующим образом.

Из рассматриваемой упорядоченной последовательности объектов выделяются упорядоченные таким же образом последовательности значений первого признака, затем второго и т.д., наконец, n-го признака. Для полученных таким образом последовательностей вычисляются значения $a^i_\xi = f_i(x^i_{j_1}, \ldots, x^i_{j_N})$, которые объединяются в строку $S_\xi = a^1_\xi, a^2_\xi, \ldots, a^n_\xi$. Из полученного списка строк выбрасываются все те строки, для которых в этом списке имеется хотя бы одна графически неотличимая. Кроме того, из списка выбрасывается каждая строка, полученная на такой последовательности объектов, для которой существует последовательность, отличающаяся только порядком следования объектов и дающая строку, графически неотличимую от какой-либо из оставшихся в списке.

Число v_k оставшихся в списке строк сопоставляется рассматриваемому множеству $\mathcal{K}_k$ и запоминается.

ШАГ 4. Из всех чисел v_k, полученных на предыдущем шаге, выбирается наименьшее, скажем v_{min}. Те множества $\mathcal{K}_k$, которым сопоставлены числа $v_k > v_{min}$, выбрасываются из дальнейшего рассмотрения. Если на этом шаге не выброшено ни одно множество $\mathcal{K}_k$ и число N меньше или равно m, то возвращаемся к шагу 3.

ШАГ 5. Все те множества G^i_k значений целевого признака для контрольного объекта, элементам которых на шаге 2 были сопоставлены множества $\mathcal{K}_k$, выброшенные на предыдущем шаге, объявляются недопустимыми, т.е. $G^i_k \notin \mathcal{E}_{on}$.

Перейдем к обсуждению универсальности, нетривиальности и содержательной инвариантности алгоритма A_0.

УТВЕРЖДЕНИЕ. Алгоритм A_0 удовлетворяет требованиям (2.1)-(2.3).

ДОКАЗАТЕЛЬСТВО. Тот факт, что алгоритм A_0 удовлетворяет требованию (2.1) (универсальность), достаточно очевиден. Требование (2.2) (нетривиальность) выполняется для алгоритма благодаря шагу 4 *).

*) Этим шагом фактически реализуется определенный критерий "семантической простоты": более простым считается такое заполнение неизвестного значения целевого признака, которое приводит к такой семантической структуре эмпирического материала, выражаемой строками $S_\xi = a^1_\xi, a^2_\xi, \ldots, a^n_\xi$, что при этом объем полученного списка строк минимален. Здесь предполагается, что предсказания, осуществляемые на основе этого критерия, будут наиболее успешными.

Покажем, что алгоритм A_0 удовлетворяет и требованию (2.3) (содержательная инвариантность). Для этого предположим, что рассматриваемая задача описана также и в других терминах, т.е. задана входная информация $\hat{\mathcal{I}}_\alpha = \langle\langle \hat{O}_\nu \rangle, \hat{K}, m, \{\hat{f}_\mu\}, \hat{\Gamma}_M \rangle$, которая связана с $\mathcal{I}_\alpha$ преобразованиями $\{\alpha^i\}$, $\{\beta^i\}$ и γ^i.

Предварительно заметим, что результаты работы алгоритма не зависят от произвола, связанного с выбором элементов из множеств G^i_k на первом шаге. Это следует из условия "в", формулируемого в определении генератора гипотез. Поэтому для простоты изложения будем предполагать, что алгоритмом A_0 как для $\mathcal{I}_\alpha$, так и для $\hat{\mathcal{I}}_\alpha$ выбираются элементы, связанные преобразованием α^i.

Проанализируем работу алгоритма A_0 на каждом шаге для входной информации $\mathcal{I}_\alpha$, а также для $\hat{\mathcal{I}}_\alpha$.

ШАГ 1. Если $\Gamma_i(\langle X^i, f_i \rangle) = \{G^i_k\}$, то $\hat{\Gamma}_M(\langle \hat{X}^M, \hat{f}_M \rangle) = \{\hat{G}^M_t\}$, причем $\{\hat{G}^M_t\} = \alpha^i[\{G^i_k\}]$.

ШАГ 2. Если A_0 для $\mathcal{I}_\alpha$ на этом шаге каждому выбранному элементу из множеств $\{G^i_k\}$ сопоставляет некоторое множество $\mathcal{K}_k$, то для $\hat{\mathcal{I}}_\alpha$ каждому выбранному элементу из множеств $\{\hat{G}^M_t\}$ на этом шаге сопоставляется множество $\hat{\mathcal{K}}_t$ такое, что $\mathcal{K}_k$ и $\hat{\mathcal{K}}_t$ связаны взаимно-однозначно преобразованиями $\{\alpha^i\}$.

ШАГ 3. Для этого шага необходимо показать, что числа v_k (сопоставляемое множеству $\mathcal{K}_k$) и $\hat{v}_t$ (сопоставляемое соответствующему множеству $\hat{\mathcal{K}}_t$) совпадают.

Заметим, что объемы исходных списков $\{S_\xi\}$ и $\{\hat{S}_\nu\}$ строк для множеств $\mathcal{K}_k$ и $\hat{\mathcal{K}}_t$ соответственно совпадают на каждом этапе образования к шагу 3 из-за одной и той же мощности этих множеств.

Очевидно, всякие две строки S_ξ и S_ζ, полученные при работе A_0 с $\mathcal{I}_\alpha$ для некоторых двух последовательностей $O_{i_1}, \ldots, O_{i_N}$ и $O_{j_1}, \ldots, O_{j_N}$ объектов из $\mathcal{K}_k$, являются графически неотличимыми тогда и только тогда, когда графически неотличимы строки $\hat{S}_\nu$ и $\hat{S}_\theta$, полученные при работе A_0 с $\hat{\mathcal{I}}_\alpha$ для таких двух последовательностей $\hat{O}_{i_1}, \ldots, \hat{O}_{i_N}$ и $\hat{O}_{k_1}, \ldots, \hat{O}_{k_N}$ объектов из $\hat{\mathcal{K}}_t$, что O_{i_1} переходит в $\hat{O}_{i_1}$, O_{i_2} – в $\hat{O}_{i_2}$ и т.д., наконец, O_{j_N} переходит в $\hat{O}_{k_N}$ в результате применения преобразований $\{\alpha^i\}$. Справедливость этого утверждения следует из определения преобразований $\{\beta^i\}$.

Из указанного вытекает, что графическая неотличимость строк, представляющая собой отношение эквивалентности на списках $\{S_\xi\}$ и

$\{\hat{s}_\nu\}$, как на множествах, порождает изоморфные модели*) $\langle \{s_\xi\}, \sigma \rangle$ и $\langle \{\hat{s}_\nu\}, \sigma \rangle$, где символом σ обозначено отношение графической неотличимости. Поэтому после первого этапа выбрасывания количество оставшихся строк как для $\{s_\xi\}$, так и для $\{\hat{s}_\nu\}$ будет одним и тем же, равным мощности фактор-множества от $\{s_\xi\}$ по σ или от $\{\hat{s}_\nu\}$ по σ.

Нетрудно показать, используя условие "в" определения семантического алгоритма, что в списке $\{s_\xi\}$ для любых классов C_1 и C_2 графически неотличимых строк, если некоторая перестановка p, примененная к некоторой последовательности объектов из $\mathcal{K}_k$, для которой получена строка $s_\xi \in C_1$, дает последовательность, которой соответствует строка $s_\zeta \in C_2$, то эта же перестановка, будучи примененной к любой последовательности объектов из $\mathcal{K}_k$, соответствующая строка которой принадлежит классу C_1, даст последовательность объектов, строка которой будет принадлежать классу C_2.

Будем говорить, что классы $C_1 \subset \{s_\xi\}$ и $C_2 \subset \{s_\xi\}$ находятся в отношении Ω, если существует перестановка объектов последовательностей, позволяющая перейти указанным образом от любого элемента класса C_1 к некоторому элементу класса C_2.

Очевидно, отношение Ω является рефлексивным (достаточно взять тождественную перестановку), симметричным (обратная перестановка) и транзитивным (суперпозиция двух перестановок), т.е. является отношением эквивалентности на множестве $\{s_\xi\}/\sigma$, представляющем собой фактор-множество от $\{s_\xi\}$ по σ.

Можно убедиться в том, что модели $\langle \{s_\xi\}/\sigma, \Omega \rangle$ и $\langle \{\hat{s}_\nu\}/\sigma, \Omega \rangle$ являются изоморфными. Отсюда следует, что количество выброшенных строк на втором этапе выбрасывания как для $\mathcal{J}_s$, так и для $\hat{\mathcal{J}}_s$ будет одинаковым, а число оставшихся строк будет совпадать с мощностью фактор-множества от множества $\{s_\xi\}/\sigma$ по Ω или, что то же самое, от множества $\{\hat{s}_\nu\}/\sigma$ по Ω. Таким образом, числа v_k и $\hat{v}_t$ совпадают.

ШАГ 4. Из анализа предыдущего шага следует, что всякое множество $\mathcal{K}_k$ будет выброшено из дальнейшего рассмотрения тогда и только тогда, когда будет выброшено соответствующее ему множество $\hat{\mathcal{K}}_t$.

ШАГ 5. Всякое множество G_k^i будет объявлено недопустимым тогда и только тогда, когда будет объявлено недопустимым множество $\hat{G}_t^\kappa$, связанное с множеством G_k^i преобразованиями $\{\alpha^i\}$, что и требовалось доказать.

*) Понятие изоморфизма моделей и в дальнейшем встречающееся понятие фактор-множества можно найти в [5].

В заключение отметим, что приведенный алгоритм, конечно же, не является единственным, для которого выполняются требования универсальности, нетривиальности и содержательной инвариантности. Рассмотрение именно этого алгоритма в настоящей работе продиктовано успешностью его модельной и практической проверки.

Литература

I. ЗАГОРУЙКО Н.Г. Методы распознавания и их применение. М., "Сов.радио", I972.

2. СУППЕС П., ЗИНЕС Дж. Основы теории измерений. -В кн.:Психологические измерения. М., "Мир", I967, с.9-II0.

3. ЗАГОРУЙКО Н.Г. Искусственный интеллект и эмпирическое предсказание. Новосибирск, НГУ, I975.

4. ВИТЯЕВ Е.Е., ГАВРИЛКО Б.П., ЗАГОРУЙКО Н.Г., САМОХВАЛОВ К.Ф. Требования к алгоритмам предсказания. -В кн.: Вычислительные системы. Вып. 50. Новосибирск, I972, с. I00-I05.

5. МАЛЬЦЕВ А.И. Алгебраические системы. М., "Наука", I970.

Поступила в ред.-изд.отд.
4 апреля I977 года