Д И С К Р Е Т Н Ы Й   А Н А Л И З

Сборник трудов

1966 г. Институт математики СО АН СССР Выпуск 7

О МАТЕМАТИЧЕСКИХ ПРИНЦИПАХ КЛАССИФИКАЦИИ ПРЕДМЕТОВ И ЯВЛЕНИЙ

А.Н.Дмитриев, Ю.И.Журавлев, Ф.П.Кренделев

Существует класс задач, в которых необходимо по ряду известных признаков предмета или явления установить, обладает ли данное явление или предмет некоторым свойством.

Так, в химии важной является задача прогноза - обладает ли данное соединение свойствами катализатора, в диагностике по ряду признаков важно установить - какой болезнью болен пациент и т.д.

В настоящей статье мы введем меру, оценивающую важность данного признака при изучении группы предметов или явлений. На основе введенной меры могут быть предложены способы классификации объектов, обладающих данным признаком, и способы отнесения нового объекта к одной из ранее изученных групп.

§ I. Допустимые таблицы

Пусть задано конечное множество $M=\{\alpha_1,\alpha_2,\ldots,\alpha_k\}$, которое в дальнейшем мы будем называть основным множеством, и на элементах M определены предикаты (свойства) $x_1, x_2, \ldots, \ldots, x_n, x_{n+1}$.

Предикат x_{n+1} назовем основным, предикаты $x_1,\ldots,x_n$ -

вспомогательными. В дальнейшем вспомогательные предикаты мы будем, если это не приводит к недоразумениям, называть просто предикатами.

Рассмотрим строки $<x_1(\tilde{\alpha}_i), x_2(\tilde{\alpha}_i), \ldots, x_n(\tilde{\alpha}_i)>$ значений вспомогательных предикатов на предмете $\tilde{\alpha}_i$, $i=1,2,\ldots,\kappa$.

Каждую строку $S_i = <x_1(\tilde{\alpha}_i), x_2(\tilde{\alpha}_i), \ldots, x_n(\tilde{\alpha}_i)>$ сопоставим с множеством строк $M(\tilde{\alpha}_i)$, которое составлено из всех строк, получающихся из S_i всевозможными заменами координат строки S_i на знак "–".

ОПРЕДЕЛЕНИЕ I. Элементы множества $M(\tilde{\alpha}_i)$ назовем э т а л о н а м и (образами) предмета $\tilde{\alpha}_i$, $i=1,2,\ldots,\kappa$.

ОПРЕДЕЛЕНИЕ 2. Таблица T, заполненная символами $\{0,1,-\}$, называется д о п у с т и м о й , если

I⁰. Существует множество $\tilde{M}$, $\tilde{M} \subset M$, $\tilde{M} = (\tilde{\alpha}_{i_1}, \ldots, \tilde{\alpha}_{i_r})$ такое, что каждая строка из T является элементом одного и только одного множества $M(\tilde{\alpha}_{ij})$, $1 \leq j \leq r$, то есть эталоном одного и только одного предмета из $\tilde{M}$.

2⁰. Все строки таблицы $T = \{a_{ij}\}$ различны. Две строки S_i и S_j таблицы T называются различными, если существует столбец – номер столбца обозначим через ℓ – такой, что

$a_{i\ell} \in \{0,1\}$, $a_{j\ell} \in \{0,1\}$ и $a_{i\ell} \neq a_{j\ell}$.

ОПРЕДЕЛЕНИЕ 3. Таблица T называется д о п у с т и м о й в узком смысле, если T является допустимой и, кроме того, на всех элементах $\tilde{M}$ (определение 2) предикат $x_{n+1} = 1$.

В дальнейшем мы будем рассматривать только допустимые или допустимые в узком смысле таблицы.

§ 2. Различающая мера признака

Пусть задана допустимая таблица T, строки которой являются образами элементов из множества $\tilde{M}$, $\tilde{M} \equiv M$, и для элементов множества $\tilde{M}$ вычислено значение основного предиката x_{n+1}. Тогда

$$\tilde{M} = \tilde{M}_1 \cup \tilde{M}_0, \quad \tilde{M}_1 \cap \tilde{M}_0 = \emptyset$$

пусто, где $\tilde{M}_1$ – множество, на элементах которого предикат $x_{n+1} = 1$ (свойство x_{n+1} выполнено), $\tilde{M}_0$ – множество, на элементах которого предикат $x_{n+1} = 0$ (свойство x_{n+1} не выполнено).

Разобьем таблицу T на две подтаблицы T_1 и T_0:

T_1 составим из всех строк таблицы T, которые являются образами элементов из $\tilde{M}_1$; T_0 – из строк, которые являются образами элементов из $\tilde{M}_0$.

ОПРЕДЕЛЕНИЕ 4. Набор $(i_1, \ldots, i_t)$ столбцов таблицы T и соответствующий ему набор $(x_{i_1}, \ldots, x_{i_t})$ назовем тестором для (T_1, T_0), если после удаления из T всех столбцов, не вошедших в число $(i_1, \ldots, i_t)$, все строки таблицы T_1 будут различны со всеми строками таблицы T_0.

ОПРЕДЕЛЕНИЕ 5. Тестор для (T_1, T_0) называется тупиковым, если после удаления из него любого столбца он перестает быть тестором для (T_1, T_0).

ПРИМЕР I. Пусть

$$T = \begin{array}{c} \begin{matrix} I & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} I & I & 0 & I & I \\ 0 & I & 0 & I & I \\ I & 0 & I & 0 & I \\ I & I & I & I & I \\ 0 & 0 & 0 & I & I \end{pmatrix} \end{array}, \quad T_1 = \begin{array}{c} \begin{matrix} I & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} I & I & 0 & I & I \\ 0 & I & 0 & I & I \\ I & 0 & I & 0 & I \end{pmatrix} \end{array}, \quad T_0 = \begin{array}{c} \begin{matrix} I & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} I & I & I & I & I \\ 0 & 0 & 0 & I & I \end{pmatrix} \end{array}.$$

Набор столбцов за номерами 2 и 3, соответственно (x_2, x_3), является тупиковым тестором для (T_1, T_0). Действительно, после удаления из T столбцов с номерами I, 4, 5 таблица T_1 переходит в T_1', T_0 в T_0' и

$$T_1' = \begin{pmatrix} I & 0 \\ I & 0 \\ 0 & I \end{pmatrix}, \qquad T_0' = \begin{pmatrix} I & I \\ 0 & 0 \end{pmatrix}.$$

Все строки таблицы T_1' отличны от всех строк таблицы T_0', следовательно, набор (x_2, x_3) является тестором для (T_1, T_0). Тупиковость (x_2, x_3) очевидна.

Пусть κ – число тупиковых тесторов таблицы T относительно (T_1, T_0) и κ_i – число тупиковых тесторов, в которые вошел столбец, соответствующий признаку x_i.

ОПРЕДЕЛЕНИЕ 6. Число $R(x_i) = \frac{\kappa_i}{\kappa}$, $i = 1,2,\ldots,n$, назовем различающим весом предиката x_i.

ПРИМЕР 2.

$$T=\begin{array}{c} 1\;2\;3\;4\;5 \\ \hline \begin{pmatrix} 0&1&0&1&0\\ 1&0&1&1&1\\ 1&0&1&1&0\\ 1&0&0&0&1\\ 0&1&0&0&0\\ 1&1&1&1&0 \end{pmatrix}\end{array},\quad T_1=\begin{array}{c} 1\;2\;3\;4\;5 \\ \hline \begin{pmatrix} 0&1&0&1&0\\ 1&0&1&1&1\\ 1&0&1&1&0 \end{pmatrix}\end{array},\quad T_0=\begin{array}{c} 1\;2\;3\;4\;5 \\ \hline \begin{pmatrix} 1&0&0&0&1\\ 0&1&0&0&0\\ 1&1&1&1&0 \end{pmatrix}\end{array}.$$

Тупиковыми тестoрами здесь являются наборы (I,2,4) и (2,3,4).

Поэтому $R(x_1)=R(x_3)=\frac{1}{2}$, $R(x_2)=R(x_4)=1$, $R(x_5)=0$.

При исследовании предмета, для которого еще не вычислен основной признак (предикат) x_{n+1} , следует в первую очередь выяснить – выполняются ли для этого предмета признаки с высоким разделяющим весом.

Вычисление разделяющего веса производится по значениям вспомогательных признаков на ранее изученных предметах,для которых значение основного признака x_{n+1} известно (или выяснено в предыдущих исследованиях).

§ 3. Алгоритмы вычисления различающего веса признаков

П е р в ы й а л г о р и т м . Введем операцию AoB под элементами множества $\{0,1,-\}$ при помощи таблицы:

A \ B	0	I	–
0	I	0	I
I	0	I	I
–	I	I	I

Тогда $S_i \circ S_j$, где $S_i=(\alpha_1,\alpha_2,\ldots,\alpha_n)$, $S_j=(\beta_1,\beta_2,\ldots,\beta_n)$, $\alpha_i\in\{0,1,-\}$, $\beta_i\in\{0,1,-\}$, $i=1,2,\ldots,n$, определяется следующим образом:

$$S_i\circ S_j=(\alpha_1\circ\beta_1,\alpha_2\circ\beta_2,\ldots,\alpha_n\circ\beta_n).$$

Пусть таблица T_1 состоит из строк $S_1,S_2,\ldots,S_\ell$, таблица T_0 – из строк $\tilde S_1,\tilde S_2,\ldots,\tilde S_m$.

Составим строки $S_{ij}=S_i\circ \tilde S_j$, $i=1,2,\ldots,\ell$, $j=1,2,\ldots,m$. В строке S_{ij} выделим номера $z_1,\ldots,z_q$ координат , равных единице, и сопоставим S_{ij} с набором $\{x_{z_1},\ldots,x_{z_q}\}=\tilde x_{ij}$. Рассмотрим множество наборов $x=\{\tilde x_{ij}\}$, $i=1,2,\ldots,m$; $j=1,2,\ldots,\ell$.

Удалим из x все наборы $\tilde x_{ij}$, для которых существуют наборы $\tilde x'_{ij}$ такие, что $\tilde x_{ij}\subset \tilde x'_{ij}$. После такой операции множество x перейдет в $x'=\{\tilde x'_{ij}\}$.

Из множества S всех наборов, составленных из некоторых букв $x_1,\ldots,x_n$, удалим все поднаборы наборов из x' .

После такой операции S перейдет в S' . Оставим в S' только наборы, все сужения которых не входят в S' . Множество всех таких наборов образует множество $\tilde T$ всех тупиковых тесторов для (T_1,T_0) . Доказательство этого факта следует из построения . Имея $\tilde T$, нетрудно построить различающие веса всех признаков $x_1,\ldots,x_n$.

ПРИМЕР 3. Рассмотрим таблицы T_1 и T_0 из примера 2 . Пусть $S_1=(01010)$, $S_2=(10111)$, $S_3=(10110)$, $\tilde S_1=(10001)$, $\tilde S_2=(01000)$, $\tilde S_3=(11110)$.

Тогда $S_{11}=(00100)$, $S_{12}=(11101)$, $S_{13}=(01011)$, $S_{21}=(11001)$, $S_{22}=(00000)$, $S_{23}=(10110)$, $S_{31}=(11000)$, $S_{32}=(00001)$, $S_{33}=(10111)$.

Поэтому $\tilde x_{11}=(x_3)$, $\tilde x_{12}=(x_1,x_2,x_3,x_5)$, $\tilde x_{13}=(x_2,x_4,x_5)$, $\tilde x_{21}=(x_1,x_2,x_5)$, $\tilde x_{22}=$ (пустой набор), $\tilde x_{23}=(x_1,x_3,x_4)$, $\tilde x_{31}=(x_1,x_2)$, $\tilde x_{32}=(x_5)$, $\tilde x_{33}=(x_1,x_3,x_4,x_5)$.

Наборы $\tilde x_{11}$, $\tilde x_{21}$, $\tilde x_{22}$, $\tilde x_{23}$, $\tilde x_{31}$, $\tilde x_{32}$ удаляем, так как они содержатся в более широких наборах. Например, $\tilde x_{21}\subset\tilde x_{12}$. Выписываем наборы, составленные из букв, содержащихся в множестве $\langle x_1,x_2,x_3,x_4,x_5\rangle$, и удаляем из них все поднаборы $\tilde x_{12}$, $\tilde x_{13}$, $\tilde x_{33}$.

В результате выполнения такой операции остаются наборы:

$$(x_1,x_2,x_4)\ ,\ (x_2,x_3,x_4)\ ,\ (x_1,x_2,x_3,x_4),$$
$$(x_1,x_2,x_4,x_5)\ ,\ (x_2,x_3,x_4,x_5)\ ,\ (x_1,x_2,x_3,x_4,x_5).$$

Из них тупиковые тесторы образуют (x_1,x_2,x_4) и (x_2,x_3,x_4). Поэтому $R(x_1)=\frac{1}{2}$, $R(x_2)=1$, $R(x_3)=\frac{1}{2}$, $R(x_4)=1$, $R(x_5)=0$.

В т о р о й а л г о р и т м . Сопоставим строкам S_{ij} , $i=1,2,\ldots,\ell$, $j=1,2,\ldots,m$, наборы $\tilde{\tilde x}_{ij}=\{x_1,\ldots,x_n\}\setminus\tilde x_{ij}$. Пусть строке S_{ij} сопоставлен набор $\{x_{ij}^1,\ldots,x_{ij}^{k(i,j)}\}$. Рассмотрим выражение:

$$\prod_{i=1}^{\ell}\prod_{j=1}^{m}(x_{ij}^1\vee\ldots\vee x_{ij}^{k(i,j)}),$$

здесь символы x_{ij}^{r} рассматриваются как булевы переменные, V - знак дизъюнкции, Π - знак конъюнкции. Полученное выражение приведем к виду $\sum\Pi$ и выполним все упрощения типа: $A \vee A = A$, $A \cdot A = A$, $A \vee AB = A$. В результате имеем

$$\vee\, x_{i_1 j_1}^{r_1} \cdot \ldots \cdot x_{i_t j_t}^{r_t} \tag{I}$$

Тогда все тупиковые тесторы для (T_1, T_0) исчерпываются наборами $(x_{i_1 j_1}^{r_1}, \ldots, x_{i_t j_t}^{r_t})$, соответствующими слагаемым в (I).

Этот алгоритм является незначительной модификацией алгоритма С.В.Яблонского [I] .

Третий алгоритм. Задача построения всех тупиковых тесторов может быть сведена к задаче расшифровки [2] специальной монотонной функции.

Сопоставим набору $(x_{i_1}, \ldots, x_{i_k})$ вершину $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_n)$ n - мерного единичного куба, положив $\alpha_{i_1} = \ldots = \alpha_{i_k} = 1$ и остальные координаты равными нулю.

Зададим функцию $F_{T_1 T_0}(x_1, x_2, \ldots, x_n)$:

$$F_{T_1 T_0}(\alpha_1 \alpha_2, \ldots, \alpha_n) = \begin{cases} 1 & \text{, если набор } (x_{i_1}, \ldots, x_{i_k}) \text{, сопоставленный вершине } (\alpha_1, \ldots, \alpha_n) \text{, образует тестор для } (T_1, T_0); \\ 0 & \text{- в противном случае.} \end{cases}$$

Функция $F_{T_1 T_0}(x_1, \ldots, x_n)$ является монотонной. Это следует из очевидных утверждений:

I^0, если набор $(x_{i_1}, \ldots, x_{i_k})$ есть тестор, то всякое расширение $(x_{i_1}, \ldots, x_{i_k})$ также образует тестор;

2^0. если набор $(x_{i_1}, \ldots, x_{i_k})$ не является тестором, то сужение наборов $(x_{i_1}, \ldots, x_{i_k})$ также не является тестором.

Очевидно, наборам, образующим тупиковые тесторы, соответствуют нижние единицы функции $F_{T_1 T_0}(x_1, \ldots, x_n)$.

Построение нижних единиц монотонной функции $F_{T_1 T_0}$ можно проводить методом В.К.Коробкова [2] .

§ 4. Отнесение эталона к классам $\tilde{M}_0$, $\tilde{M}_1$

Пусть заданы таблицы T_1 , T_0 эталонов определенных предметов, у которых основной признак равен соответственно единице, нулю.

Задан допустимый эталон S , для которого значение основного признака неизвестно и вычисление этого значения затруднено.

Тогда в ряде случаев для отнесения S к T_1 или к T_0 можно пользоваться следующей процедурой.

Пусть

$$T_1 = \begin{pmatrix} \alpha_{11} \alpha_{12} \ldots \alpha_{1n} \\ \alpha_{21} \alpha_{22} \ldots \alpha_{2n} \\ \ldots\ldots\ldots \\ \alpha_{m1} \alpha_{m2} \ldots \alpha_{mn} \end{pmatrix}, \quad T_0 = \begin{pmatrix} \beta_{11} \beta_{12} \cdots \beta_{1n} \\ \beta_{21} \beta_{22} \cdots \beta_{2n} \\ \ldots\ldots\ldots \\ \beta_{\ell 1} \beta_{\ell 2} \cdots \beta_{\ell n} \end{pmatrix}, \quad S = (\gamma_1, \ldots, \gamma_n).$$

Для каждого признака $x_1, \ldots, x_n$ вычисляется разделяющий вес $R(x_i)$, $i = 1, 2, \ldots, n$.

Вычисляется значение величины:

$$\rho_1 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \overline{(\alpha_{ij} \circ \gamma_j)}}{m}, \qquad \rho_2 = \frac{\sum_{i=1}^{\ell} \sum_{j=1}^{n} \overline{(\beta_{ij} \circ \gamma_j)}}{\ell}.$$

Если $\rho_1 > \rho_2$, делается прогноз: основной признак для предмета, которому соответствует эталон S , не выполнен. Если $\rho_1 < \rho_2$, то делается прогноз: основной признак для предмета, которому соответствует эталон S , выполнен.

Области применимости и точность описанного метода будут исследованы в дальнейших публикациях.

§ 5. Информационный вес признака

Пусть задана допустимая в узком смысле таблица T (см. определение 3), то есть на всех предметах, эталонами которых являются строки таблицы T , основное свойство x_{n+1} выполнено.

При решении различных задач бывает необходимо расклассифицировать вспомогательные признаки $x_1, \ldots, x_n$ по степени ценности их для изучения свойства x_{n+1} , иными словами, необходимо установить последовательность выявления признаков для оценки явления.

Ниже мы введем меру ценности (информационный вес) признака x_i .

ОПРЕДЕЛЕНИЕ 4$'$. Набор $(i_1, \ldots, i_\ell)$, соответственно $(x_{i_1}, \ldots, x_{i_\ell})$, называется тестом таблицы T , если после удаления из T всех столбцов, за исключением $(i_1, \ldots, i_\ell)$, получается таблица, все строки которой различны.

ОПРЕДЕЛЕНИЕ 5. Тест $(i_1,\dots,i_\ell)$, соответственно $(x_{i_1},\dots,x_{i_\ell})$, называется тупиковым для таблицы T, если из него нельзя удалить ни одного столбца без того, чтобы он перестал быть тестом.

Понятие "тест" впервые было введено С.В.Яблонским в связи с изучением методов контроля электрических схем [I].

Пусть κ – число тупиковых тестов таблицы T, κ_i – число тупиковых тестов таблицы T, в которое входит столбец, соответствующий признаку (предикату) x_i.

ОПРЕДЕЛЕНИЕ 7. Число $\rho(i)=\frac{\kappa_i}{\kappa}$ называется информационным весом признака x_i, $i=1,2,\dots,n$.

Величина $\rho(i)$ оценивает важность признака x_i при изучении явления, описываемого основным предикатом x_{n+1}. Если расположить признаки x_i, $i=1,2,\dots,n$, в последовательность по убыванию $\rho(i)$, то признаки, получившие меньший номер, оказываются более важными при изучении предметов, обладающих свойством x_{n+1}.

При исследовании конкретных таблиц T, описывающих естественнонаучные эксперименты, обычно оказывается, что признаки распадаются по значениям $\rho(i)$ на группы, причем колебание между группами величины $\rho(i)$ велико по сравнению с колебанием внутри группы. Таким образом, признаки по информационной ценности часто распадаются на четко очерченные ранги. Признаки первого ранга оказываются наиболее существенными при изучении предметов, обладающих свойством x_{n+1}, второго ранга – менее существенными и т.д.

ЗАМЕЧАНИЕ. Таблицы, рассматриваемые в данном параграфе, должны обладать еще одним свойством ; в каждом столбце таблицы имеется хотя бы один элемент, равный нулю, и хотя бы один элемент, равный единице.

ПРИМЕР 4. (иллюстрированный). В качестве основного множества M рассматривается совокупность книг. Основной признак: книга $\tilde{\alpha}$ является книгой для детей.

Вспомогательные признаки:

$x_1(\tilde{\alpha})$ – $\tilde{\alpha}$ набрана крупным шрифтом ;
$x_2(\tilde{\alpha})$ – в $\tilde{\alpha}$ много иллюстраций ;
$x_3(\tilde{\alpha})$ – $\tilde{\alpha}$ издана на хорошей бумаге ;
$x_4(\tilde{\alpha})$ – в $\tilde{\alpha}$ помещено хотя бы одно стихотворение ;
$x_5(\tilde{\alpha})$ – сюжет $\tilde{\alpha}$ не представляет собой единого целого.

Рассматривается множество M, состоящее из шести книг

для детей (на элементах M основной предикат выполнен).

Таблица T соответствующих эталонов:

	x_1	x_2	x_3	x_4	x_5
S_1	I	0	0	I	I
S_2	I	I	0	I	I
S_3	I	0	I	I	0
S_4	I	I	I	0	0
S_5	0	I	I	0	I
S_6	0	I	I	I	0

Тупиковые тесты таблицы T:

$$\langle x_1,x_2,x_5\rangle,\ \langle x_2,x_4,x_5\rangle,\ \langle x_1,x_2,x_3,x_4\rangle.$$

Откуда получаем возможность вычислить $\rho(i)$:

$$\rho(1)=\frac{1}{2}\ \frac{2}{3},\ \rho(2)=1,\ \rho(3)=\frac{1}{3},\ \rho(4)=\frac{1}{2}\ \frac{2}{3},\ \rho(5)=\frac{1}{2}\ \frac{2}{3}$$

Наиболее важным для характеристики свойства книги быть детской оказывается признак $x_2(\tilde{\alpha})$ в $\tilde{\alpha}$ много иллюстраций, наименее важным – признак $x_3(\tilde{\alpha})$, характеризующий качество бумаги в $\tilde{\alpha}$. По значениям $\rho(i)$ признаки распадаются на три группы:

$\{x_2\}$ – группа первого ранга, $\rho(2)=1$;
$\{x_1,x_4,x_5\}$ – группа второго ранга, $\rho(1)=\rho(4)=\rho(5)=\frac{1}{2}\ \frac{2}{3}$;
$\{x_3\}$ – группа третьего ранга, $\rho(3)=\frac{1}{3}$.

ПРИМЕР 5. При решении задачи, возникшей в одной из областей естествознания, была получена группа семи эталонов, для которых основной признак (принадлежность к определенному типу) был выполнен. Рассматривались вспомогательные признаки : $x_1, x_2, \dots, x_{17}$.

Таблица T эталонов имеет следующий вид:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}
I	0	0	I	I	I	I	0	I	I	I	0	I	I	I	I	I
I	0	0	I	0	0	I	I	0	I	I	0	I	I	I	I	0
0	0	I	0	0	0	I	I	0	I	I	0	I	I	0	I	I
0	I	0	0	0	0	I	I	0	0	0	I	I	I	I	I	–
I	0	0	0	0	0	I	I	0	0	0	I	I	0	I	I	–
0	0	I	–	0	0	I	–	–	0	0	I	0	0	I	I	–
I	0	0	I	–	0	0	0	I	0	0	I	I	0	I	I	–

Оказалось, что таблица T характеризуется 406 тупиковыми тестами, для которых вычислено:

Признаки	x_{16}	x_7	x_{13}	x_2	x_4	x_1	x_{14}	x_{17}	x_6	x_{15}	x_3
Число тупиковых тестов, в которые входит данный признак	374	334	322	3I6	256	250	243	242	228	225	2II
Информационный вес признака	097	087	083	082	066	065	063	063	059	058	054

Признаки	x_5	x_{12}	x_{10}	x_{11}	x_8	x_9
Число тупиковых тестов, в которые входит данный признак	I55	I53	I52	I5I	I2I	II9
Информационный вес признака	040	040	039	039	032	030

Таким образом, признаки распались на группы:

$\{x_{16}\}$ – группа первого ранга, $\rho(16)=0,97$;

$\{x_7, x_{13}, x_2\}$ – группа второго ранга, $0,82\leqslant\rho(i)\leqslant 0,87$;

$\{x_4,x_1,x_{14},x_{17},x_6,x_{15},x_3\}$ – группа третьего ранга, $0,54\leqslant\rho(i)\leqslant 0,66$;

$\{x_5, x_{12}, x_{10}, x_{11}\}$ – группа четвертого ранга, $0,39\leqslant\rho(i)\leqslant 0,40$;

$\{x_8, x_9\}$ – группа пятого ранга, $0,30\leqslant\rho(i)\leqslant 0,32$.

Оказалось, что упорядоченность признаков по убыванию информационного веса соответствует упорядоченности по важности (в содержательном смысле).
Так, признаки x_{16}, x_7, x_{13}, x_2 наиболее характерны для данного типа эталонов.

§ 6. Вычисление информационного веса признака

Вычисление информационного веса $\rho(i)$ сводится к построению всех тупиковых тестов таблицы эталонов T .

Алгоритмы выделения всех тупиковых тестов имеют ту же природу, что и алгоритмы синтеза всех тупиковых тесторов для (T_1, T_0), описанные в § 3.

Рассмотрим аналоги второго и третьего алгоритмов из § 3.

I) Алгоритм С.В.Яблонского.

Пусть таблица T состоит из строк $S_1, S_2, \ldots, S_t$. Для каждой пары i, j , $i\neq j$, составляем строку $S_{ij}=S_i \circ S_j$ и выделяем набор x_{ij} переменных $x'_{ij}, \ldots, x_{ij}^{\kappa(i,j)}$, для которых соответствующая координата в S_{ij} равна нулю.

Составляем выражение:

$$\prod_{\substack{i,j \\ 1\leqslant i\leqslant t \\ 1\leqslant j\leqslant t \\ i\neq j}} (x'_{ij} \vee \ldots \vee x_{ij}^{\kappa(i,j)}). \qquad (2)$$

Рассматривая символы x_{ij}^{z} как булевы переменные, приводим (2) к виду $\sum\prod$ и производим все упрощения $A\cdot A=A$, $A\vee A=A$, $A\vee A\cdot B=A$. Пусть (2) переходит в

$$\vee x'_{i_1 j_1} \cdots x^{z}_{i_z j_z}. \qquad (3)$$

Каждому слагаемому в (3) соответствует тупиковый тест $(x'_{i_1 j_1}, \ldots, x^{z}_{i_z j_z})$, и все тупиковые тесты получаются таким образом.

2) Сведение синтеза всех тупиковых тесторов к задаче расшифровки монотонной функции. Сопоставим набору $(x_{i_1}, \ldots, x_{i_k})$ вершину $\tilde{\alpha}=(\alpha_1, \ldots, \alpha_n)$, положив $\alpha_{i_1}=\ldots=\alpha_{i_k}=1$ и остальные координаты равными нулю.

Зададим функцию $f_T(x_1, \ldots, x_n)$.

$$f_T(\alpha_1,\ldots,\alpha_n)=\begin{cases} 1, & \text{если набор } (x_{i_1},\ldots,x_{i_k}), \text{ сопоставленный вершине } (\alpha_1,\ldots,\alpha_n), \text{ образует тест для } T ; \\ 0 & \text{– в противном случае .}\end{cases}$$

ПРИМЕР 6. Пусть

$$T=\begin{pmatrix} 0 & 0 & 0 & 0 \\ I & 0 & I & 0 \\ 0 & I & I & I \\ I & I & I & I \end{pmatrix}.$$

Тогда функция $f_T(x_1, x_2, x_3, x_4)$ несущественно зависит от переменной x_3 ;

$$f_T(x_1, x_2, x_3, x_4)=\varphi(x_1, x_2, x_4)=x_1\cdot x_2 \vee x_1\cdot x_4 .$$

§ 7. Информационный вес эталона (строки)

Пусть информационный вес вспомогательных предикатов оказался равным $\rho(1), \ldots, \rho(n)$.

Рассмотрим произвольный эталон S (строку таблицы T). Пусть в строке S на местах $i_1,\dots,i_e$ стоят единицы, на местах $j_1,\dots,j_r$ – прочерки. Обозначим через q_j^t число элементов, отличных от прочерка, в столбце j_t, $t=1,2,\dots,r$; через $\tilde{q}_j^t$ – число элементов (равных единице) в столбце j_t, $t=1,2,\dots,r$.

Обозначим:

$$\frac{\tilde{q}_j^t}{q_j^t}=q(j_t),\ t=1,2,\dots,r.$$

$$\sum_{i=1}^{n} p(i)=p.$$

ОПРЕДЕЛЕНИЕ 8. Величина $\mathcal{I}(S)=\frac{1}{p}\left(\sum_{j=1}^{\ell} p(i_j)+\sum_{t=1}^{r} q(j_t)\cdot p(j_t)\right)$ называется информационным весом эталона S.

Если число эталонов в таблице T невелико, вместо величины $\mathcal{I}(S)$ можно рассматривать более грубую характеристику:

$$\tilde{\mathcal{I}}(S)=\frac{1}{p}\left(\sum_{j=1}^{\ell} p(i_j)+\frac{1}{2}\sum_{t=1}^{r} p(j_t)\right).$$

В приложениях часто возникают задачи классификации эталонов (предметов), для которых выполнен предикат x_{n+1} по степени проявления свойства x_{n+1} (степени активности вещества, степени опасности данной инфекции и т.д.).

Оказалось, что в ряде задач упорядоченность предметов по убыванию величины $\mathcal{I}(S)$ совпадает (или сильно коррелирована) с упорядоченностью по степени проявления свойства x_{n+1}. Таким образом, значения $\mathcal{I}(S)$ могут послужить основой для построения классификации предметов или явлений по степени проявления свойства x_{n+1}.

ПРИМЕР 7. В таблице примера 4 имеем:

$$p=\frac{17}{6};\quad \mathcal{I}(S_1)=\frac{6}{17}\left(p(1)+p(4)+p(5)\right)=\frac{9}{17};\ \mathcal{I}(S_2)=\frac{15}{17};$$

$$\mathcal{I}(S_3)=\frac{8}{17};\quad \mathcal{I}(S_4)=\mathcal{I}(S_5)=\mathcal{I}(S_6)=\frac{11}{17}.$$

Таким образом, среди книг, эталонами которых являются S_1, S_2, S_3, S_4, S_5, S_6, наиболее типичная детская книга представлена эталоном S_3.

Характеристика S_3: книга набрана крупным шрифтом, в ней

много иллюстраций, качество бумаги – невысокое, в книге есть стихи, содержание книги не представляет собой единого целого.

Литература

1. И.А.Чегис, С.В.Яблонский. Логические способы контроля электрических схем. – Труды Математического института им. В.А.Стеклова, 1958, т. 51. стр.270–360.

2. В.К.Коробков. О монотонных функциях алгебры логики. – Проблемы кибернетики, 1965, вып. 13, стр. 5.

Поступила в редакцию
8.IV.1966 г.