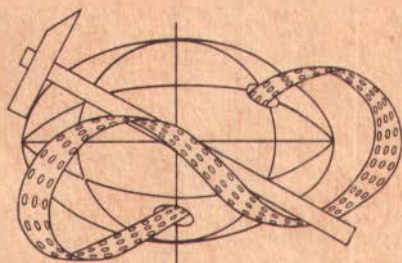


АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ГЕОЛОГИИ И ГЕОФИЗИКИ

ЛОГИКО-МАТЕМАТИЧЕСКАЯ ОБРАБОТКА ГЕОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

(ОПЕРАТИВНО-ИНФОРМАЦИОННЫЙ МАТЕРИАЛ)



НОВОСИБИРСК-1975

АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ГЕОЛОГИИ И ГЕОФИЗИКИ

**ЛОГИКО-МАТЕМАТИЧЕСКАЯ
ОБРАБОТКА
ГЕОЛОГИЧЕСКОЙ
ИНФОРМАЦИИ**

**ПРОГРАММЫ К ЭВМ
ДЛЯ ЛОГИКО-МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ
ГЕОЛОГИЧЕСКОЙ ИНФОРМАЦИИ**

(ОПЕРАТИВНО-ИНФОРМАЦИОННЫЙ МАТЕРИАЛ)

НОВОСИБИРСК—1975

А Н Н О Т А Ц И Я

В сборнике представлены программы, реализующие алгоритмы логико-математической обработки геологической информации для решения задач, в основном, прогнозно-поискового профиля, а также ряд программ для обработки статистических данных.

Программы ориентированы на машины М-220, м-222, БЭСМ-6.

Сборник рассчитан на широкий круг научных и инженерно-технических работников, интересующихся вопросами математической обработки геологической информации.

Редакционная коллегия

В.О.Красавчиков, Е.А.Смертин, Т.И.Штатнова.

Ответственный редактор А.Н.Дмитриев

Авторский коллектив:

А.Н.Дмитриев, В.В.Бабич, Е.И.Бурова, В.Н.Кандыба,
В.Д.Карбышев, В.О.Красавчиков, М.А.Левчук, Ю.Я.Лыков,
А.П.Прокопенко, Т.Л.Слущкая, Е.А.Смертин, А.Д.Соколов,
Г.С.Федосеев, Т.И.Штатнова.

Печатается по решению секции
стратиграфии, литологии и осадочных
полезных ископаемых Ученого Совета
Института геологии и геофизики СО АН СССР

© Институт геологии
и геофизики СО АН СССР,
1975 г.

В настоящем сборнике публикуются алгоритмы и программы, предназначенные для обработки информации при решении геологических задач, в основном, прогнозно-поискового профиля. Главным образом это программы, реализующие составные части логико-математического подхода к решению практических задач. Большинство из них разработано сотрудниками лаборатории логико-математической обработки геологической информации Института геологии и геофизики. Ряд программ написан в Институте математики СО АН СССР. Сборник включает 18 программ, каждая из которых сопровождается инструкцией к пользованию и контрольным примером с выдачей результатов на печать.

Программы составлены для ЭВМ БЭСМ-6, М-220, М-222 в машинных кодах (ручные), на языках АЛГОЛ- α , АЛГОЛ- α -6, АЛГОЛ-БЭСМ. Они сгруппированы в 4 типа по виду и способу получения информационных оценок. Соответственно и сам сборник состоит из четырех частей.

К первому типу (часть I) относятся алгоритмы и программы, основанные на применении спектрального метода оценки столбцов и строк числовых матриц (метода согласованных оценок), предложенного Ю.Л.Васильевым и А.Н.Дмитриевым ("Качели"). Во второй части собраны программы тестового подхода к классификации предметов и явлений, основанного на понятии теста С.В.Яблонского и развитого в работах А.Н.Дмитриева, Ю.И.Журавлева, Ф.П.Кренделева, Р.М.Константинова и др. Третий тип программ (часть III) — это программы, реализующие разработанный В.В.Бабичем, В.О.Красавчиковым и Г.С.Федосеевым метод суммарного учета мер приуроченности и согласования (связи) косвенного и целевого признаков ("Каскад"). Четвертый тип объединяет программы статистического характера.

Основой программ I типа (3 программы) является итерационная процедура спектрального подхода. Программы П1 и П3 предназначены для вычисления нагрузок строк и столбцов бинарной таблицы и вычисляемые по ним нагрузки отличаются только коэффициентами пропорциональности (способом нормировки нагрузок). Программа П2 используется для расчета коэф-

фициентов, которые применяются при распознавании в двухклассовой постановке.

Второй тип программ, в основе которых лежит известный тестовый метод, представлен семью программами. Программы П1 и П2 предназначены для вычисления тестовых информационных весов столбцов бинарных таблиц. Кроме того, с помощью П1 вычисляются информационные веса строк, а также, при соответствующей модификации, Q-тестовые веса столбцов и строк. Программа П3 используется при вычислении тестовых, Q-тестовых и пакетных информационных весов столбцов. Программа П4 применяется для проведения голосования по тестам. С помощью программы П5 производится голосование по тесторам, другие применения этой программы – голосование по Q-тестам и голосование по пакетам. Программой П6 определяются количества тупиковых тесторов всех имеющихся длин и другие параметры. По программе П7 вычисляются тестовые и Q-тестовые информационные веса столбцов и строк бинарных 5-строчных таблиц с любым числом столбцов.

Общей особенностью I и II типов программ является то, что в процессе вычисления нагрузок столбцов и строк учитывается "сцепленность" между собой характеристических признаков, причем целевой признак используется для формирования классов, подлежащих обработке, и не влияет непосредственно на вычисление информационных оценок сформированных таблиц.

Алгоритмы, составляющие третий тип, содержат 5 программ и отличаются тем, что для вычисления оценок столбцов и строк привлекают непосредственно в обработку значения целевого признака, а само это вычисление построено на индивидуальной оценке каждого отдельного признака. Применение формулируемых методов оценки признаков позволяет проводить вычислительные процедуры в малое время и практически не ограничивать себя в количестве признаков или объектов исследования. Программы П1 и П2 предусматривают вычисление оценок информативности нескольких видов для столбцов бинарных таблиц. Логическим продолжением программ П1 и П2 является программа П3. Она предназначена для получения оценок строк согласно заданным оценкам столбцов и позволяет проводить анализ однородности эталонных выборок и распознавание объектов экзамена и проб. Кроме того, эта программа ранжирует столбцы по убыванию оценок и выдает на печать все данные, необходимые для мини-

мизации признакового пространства.

Программа П4 представляет собой расширенный вариант программ П1 и П2. Она позволяет выбирать оптимальные варианты бинарного кодирования количественной и порядковой информации в соответствии с целевым заданием. Одновременно вычисляются требуемые условиями задачи оценки столбцов и выдается на печать исходная таблица, закодированная оптимальным образом в бинарном коде.

Программа П5 предназначена для оптимального бинарного кодирования признаков. После кодирования производится подсчет нагрузок строк и минимизация пространства признаков с учетом "идеальных" объектов. Программу П6 можно использовать для 2-классовых постановок задач.

в четвертую часть вошли 3 программы. В их основу положены известные методы математической статистики: метод главных компонент, вычисление корреляционных отношений и аппроксимации кривых линий.

Программа П1 предназначена для вычисления главных компонент (факторов) и факторных нагрузок (собственных векторов) на основе анализа ковариационной и корреляционной матриц и получения новых значений (координат) для каждого исходного объекта на основе полученных факторов и факторных нагрузок.

С помощью программы П2 производится вычисление корреляционных отношений для анализа линейных и нелинейных корреляционных связей между геологическими признаками. Кроме того, в программе П2 однозначно реализована группировка значений геологических признаков по формуле Стерджесса, что дает возможность получать менее субъективные величины корреляционных отношений.

Программа аппроксимации кривых П3 предназначена для получения параметров радиуса кривизны и центров мгновенных радиусов кривизны.

Следует отметить, что все описанные программы апробированы при решении практических прогностико-поисковых и диагностических задач. В процессе применения описанных алгоритмов уточнялись области их применения и путем сравнительного анализа выяснялись возможности их комплексного использования. Для удобства изложения все необходимые сведения общематематического характера (определения, описания алгоритмов, формулировки теорем и т.д.) для каждого типа программ сведены в одно целое в начале каждого типа. По сравнению с общепринятым уровнем справочный математический материал несколько расширен с той целью, чтобы по возможности облегчить эксплуатацию программ специалистами-нематематиками.

ПРОГРАММЫ МЕТОДА СОГЛАСОВАННЫХ ОЦЕНОК

В настоящей части опубликованы программы для оценки строк и столбцов таблиц описаний по методу согласованных оценок [I] и программы для процедур распознавания, основанных на этих оценках. Изложим основы метода согласованных оценок.^{х)}

Пусть имеется h объектов W_1, \dots, W_h , охарактеризованных двузначными признаками P_1, \dots, P_l . Пусть $T = (t_{ij})$ - таблица размера $h \times l$, которая составлена из единиц и нулей, и в которой i -я строка (t_{i1}, \dots, t_{il}) отвечает объекту W_i , $i=1, \dots, h$, j -й столбец (t_{1j}, \dots, t_{hj}) отвечает признаку P_j , $j=1, \dots, l$. Таблица T отражает выраженность признаков у объектов: если для i -го объекта и j -го признака она превышает некоторый уровень, то $t_{ij} = 1$, а если меньше, то $t_{ij} = 0$. Условимся, что $h \geq 2$, $l \geq 2$ и что в таблице T нет строк, и нет столбцов, составленных сплошь из нулей.

Если признаки k -значны, $k > 2$, то t_{ij} может принимать соответствующее число значений из отрезка $[0, 1]$. Ограничимся рассмотрением случая $k=2$, так как переход на случай $k > 2$ будет очевиден.

Вектор $\bar{a} = (a_1, \dots, a_m)$ называется положительным, если $a_1, \dots, a_m > 0$. Норма вектора $\bar{a} = (a_1, \dots, a_m)$ - число $|\bar{a}|$, равное $\prod_{i=1}^m a_i$; вектор \bar{a} называется нормальным, если $|\bar{a}| = 1$.

Числовую меру для объектов признаков естественно задавать в виде положительных нормированных векторов $\bar{\omega} = (\omega_1, \dots, \omega_h)$ и $\bar{\pi} = (\pi_1, \dots, \pi_l)$. Ниже мы определим их как нагрузку строк и столбцов таблицы T . По ним предлагается судить о проявленности фактора $X_{n,i}$ соответственно в объектах W_1, \dots, W_h а также о степени влияния на нее признаков P_1, \dots, P_l .

^{х)}Ниже даны принадлежащие Ю.Л.Васильеву разделы 3,4,5 статьи [I]

Определение нагрузки подсказывается анализом грубой оценки объектов и признаков, при которой им приписывается вес, равный количеству единиц в соответствующих строках и столбцах таблицы Т. При этом строки и столбцы выступают как бы порознь.

	P_1	P_2	P_3	P_4	Вес
W_1	I	I	0	0	2
W_2	0	I	I	0	2
W_3	0	0	I	I	2
Вес	I	2	2	I	

Рис.1 Грубая оценка

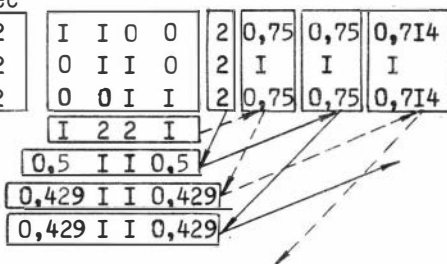


Рис.2 Итерационный процесс ("Качели")

В примере на рис.1 строки получают равные веса, однако они различны по отношению к столбцам: во второй строке обе единицы отвечают признакам, представленным единицами и в других строках, а в первой и третьей строках имеются лишь по одной такого рода единице. Содержательно это может означать, что вторая строка отвечает более сильному проявлению фактора X_{n+1} .

Определение нагрузки отражает итог описанных ниже пересчетов весов строк с учетом весов столбцов, и наоборот. При каждом пересчете вес строки $W_i, i=1, \dots, h$ определяется через веса столбцов, найденные на предшествующем шаге как сумма весов тех столбцов, по которым в строке W_i стоят единицы; вес столбца $P_j, j=1, \dots, l$ аналогично определяется через найденные на предшествующем шаге веса строк; затем полученные два набора чисел нормируются, начинается новый пересчет и т.д. Процесс сходится к некоторым предельным векторам \bar{W} и \bar{P} , которые примем в качестве нагрузок. Для таблицы на рис. 1 $\bar{W} \approx (0,707; I; 0,707)$, $\bar{P} \approx (0,414; I; I; 0,414)$. По сравнению с грубой оценкой нагрузки \bar{W} выделяет вторую оценку. На рис.2 представлены результаты трех первых пересчетов.

Исходя из таблицы Т и пары векторов $\bar{W} = (\omega_1, \dots, \omega_h)$, $\bar{P} = (P_1, \dots, P_l)$ определим пару векторов $\bar{W}' = (\omega'_1, \dots, \omega'_h)$

$\bar{P} = (\bar{P}_1, \dots, \bar{P}_l)$, а также нормирующие множители α, β и пару векторов \bar{W} и \bar{P}' .

$$\bar{W} = T(\bar{P}), \text{ т.е. } \omega_i = t_{i1}p_1 + \dots + t_{il}p_l, \bar{W}' = \alpha \cdot \bar{W}, \quad (1)$$

$$i = 1, \dots, h, \quad |\bar{W}'| = 1$$

$$\bar{P}' = T^*(\bar{P}), \text{ т.е. } \bar{P}'_j = t_{1j}\omega_1 + \dots + t_{hj}\omega_h, \bar{P}' = \beta \cdot \bar{P}, \quad (2)$$

$$j = 1, \dots, l, \quad |\bar{P}'| = 1$$

Формулы (1), (2) дают формальное определение упомянутых пересчетов.

О п р е д е л е н и я . Пару нормированных векторов $\bar{\varphi} = (\varphi_1, \dots, \varphi_h)$, $\bar{\psi} = (\psi_1, \dots, \psi_l)$ назовем ф р а к ц и о н и р у ю щ е й п а р о й т а б л и ц ы , а сами векторы — ф р а к ц и о н и р у ю щ и м и в е к т о р а м и , если $\bar{\varphi}' = \varphi$ и $\bar{\psi}' = \psi$.

Фракционирующую пару векторов $\bar{\omega}, \bar{\pi}$ назовем н а г р у з - к о й т а б л и ц ы T , а сами векторы — н а г р у з к о й с т р о к и н а г р у з к о й с т о л б ц о в , если эти векторы положительны.

Таблицу T, нагрузка которой существует и единственна, назовем и з м е р и м о й .

Таблицу T назовем р а з р о з н е н н о й , если среди всех ее h строк найдутся h_1 таких строк, каждая из которых ортогональна^{*)} каждой из прочих $h - h_1$ строк ($h_1 < h$). Иными словами, совокупность объектов распадается на две такие группы по h_1 и $h - h_1$ объектов, что каждый из признаков обязательно представлен нулями во всех объектах какой-то одной из этих двух групп (грубо говоря, эти две группы "не имеют ничего общего"). Перестановками строк и столбцов любую разрозненную таблицу можно привести к виду $\begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix}$. Любая разрозненная

таблица неизмерима^{*)}

*) Можно доказать, что доля разрозненных таблиц среди всех $(h \times l)$ таблиц T стремится к нулю при $h \leq l \leq 2^{h-1}$ и $h \rightarrow \infty$

**) Два вектора (x_1, \dots, x_l) и (y_1, \dots, y_l) ортогональны, если $x_1y_1 + \dots + x_ly_l = 0$

Таблицу T , не являющуюся разрозненной, назовем- с в я з н о й. Любая связанная таблица измерима.

Нагрузка строк $\bar{\omega}$ и нагрузка столбцов $\bar{\pi}$ измеримой таблицы T являются пределами соответственно последовательностей векторов

$$\bar{\omega}^0, \bar{\omega}^1, \dots, \bar{\omega}^n, \dots,$$

$$\text{где } \bar{\omega}^0 = \overbrace{(1, \dots, 1)}^{h \text{ единиц}}, \bar{\omega}^n = (\bar{\omega}^{n-1})', \quad (3)$$

$$\bar{\pi}^0, \bar{\pi}^1, \bar{\pi}^2, \dots, \bar{\pi}^n, \dots,$$

$$\text{где } \bar{\pi}^0 = \overbrace{(1, \dots, 1)}^{l \text{ единиц}}, \bar{\pi}^n = (\bar{\pi}^{n-1})', \quad (4)$$

На основе этой же процедуры может быть произведена обработка таблиц описаний, составленных из произвольных неотрицательных чисел.

Кроме того, метод согласованных оценок может быть использован для обработки двузначных таблиц, составленных из чисел $(0,1)$ или $(1,-1)$. Следует отметить, что в том случае, когда таблица T составлена из единиц и минус единиц, то для подсчета нагрузок столбцов и строк в качестве нормы вектора

$$\bar{a} = (a_1, a_2, \dots, a_m)$$

берется число $|\bar{a}| = \sum_{i=1}^m |a_i|$.

Здесь следует использовать программу П 2. На основе метода согласованных оценок может быть сформулирован ряд процедур распознавания объектов, охарактеризованных бинарными признаками. Описываемый в настоящей работе алгоритм распознавания, использующий качельные нагрузки столбцов и строк "типизированных" таблиц эталонов, предложен В.О.Красавчиковым.

Пусть имеются две группы объектов

$$S_1^1, \dots, S_{m_1}^1 \quad \text{и} \quad S_1^2, \dots, S_{m_2}^2$$

Из их описаний обычным путем составим таблицы T_1 и T_2 . Строка

$$(\tilde{t}_1^k, \dots, \tilde{t}_n^k) = \tilde{s}^k,$$

где $k=1,2$, называется типичной строкой таблицы T_k , если:

$$\tilde{t}_j^k = \begin{cases} 1 & \text{при } \sum_{i=1}^{m_k} t_{ij}^k \geq \frac{m_k}{2} \\ 0 & \text{при } \sum_{i=1}^{m_k} t_{ij}^k < \frac{m_k}{2} \end{cases}, \quad \text{где } t_{ij}^k \text{ - элемент таблицы } T_k^{\#}.$$

Рассмотрим следующую достаточно общую схему распознавания, использующую понятие типичной строки ("реплики" [5]). Пусть $P^k = (P_1^k, \dots, P_n^k)$ произвольный неотрицательный вектор нагрузок столбцов таблицы T_k , $k=1, 2, P_j^k \geq 0, j=1, \dots, n, \sum_{j=1}^n P_j^k = 1$.

Для произвольной строки $S = (t_1, \dots, t_n)$ определим

$$z(P^k, S) = \sum_{j=1}^n |t_j + \tilde{t}_j^k - 1| \cdot P_j^k$$

$$R(P^1, P^2, S) = \frac{z(P^1, S)}{z(P^2, S)} \quad (z(P^2, S) \neq 0)$$

Величина $z(P^k, S)$ представляет собой взвешенное число совпадений строки S с типичной строкой таблицы T_k , т.е. меру близости S к классу, представленному группой объектов $S_{1k}, \dots, S_{m_k k}$.

Величина $R(P^1, P^2, S)$ оценивает тяготение строки S к одному из указанных двух классов при заданных мерах близости. Если $R(P^1, P^2, S) > 1$, то S тяготеет к первому классу, если $R(P^1, P^2, S) < 1$, то ко второму. При $R(P^1, P^2, S) = 1$ мы ничего не можем сказать о принадлежности S . Естественно сформулировать такое решающее правило для диагностики испытуемых объектов S_n , описаниями которых являются строки $S = (t_1, \dots, t_n)$:

- а) при $R(P^1, P^2, S) \geq 1 + \varepsilon_1$ S_n относится к первому классу;
- б) при $R(P^1, P^2, S) \leq 1 - \varepsilon_2$ S_n относится ко второму классу;
- в) при $1 - \varepsilon_2 < R(P^1, P^2, S) < 1 + \varepsilon_1$ S_n не распознается.

Здесь пороги $\varepsilon_1, \varepsilon_2 > 0$ определяют "решительность" алгоритма распознавания.

Наиболее простой процедурой подобного рода является, по-видимому, следующая: положим $P = P_{\bar{n}}^k = (\frac{1}{n}, \dots, \frac{1}{n})$ т.е. все признаки считаются в равной степени существенными для диагностики. Тогда $z(P_{\bar{n}}^k, S) = z_{\bar{n}}(S) = \frac{1}{n} \sum_{j=1}^n |t_j + \tilde{t}_j^k - 1|$, т.е. равняется числу совпадений строки S с типичной строкой \tilde{S}^k поделенному на n , а $R(P_{\bar{n}}^1, P_{\bar{n}}^2, S) = R_{\bar{n}}(S)$ - отношение числа совпадений с типичной строкой первого класса к числу совпадений с типичной строкой второго класса.

Для того, чтобы указанное решающее правило можно было приме-

* При $\sum_{i=1}^{m_k} t_{ij}^k = \frac{m_k}{2}$ значение \tilde{t}_j^k может определяться или из неформальных, практических соображений, или случайным путем.

нять к решению той или иной конкретной задачи диагноза, необходимо убедиться в его способности распознавать объекты обучения $S_1^1, \dots, S_{m_1}^1, S_1^2, \dots, S_{m_2}^2$, т.е. проверить, что если S_j^k — строка соответствующая объекту S_j^k , где $j=1, \dots, m_k$, то

$$(*) \quad \begin{cases} R(\rho^1, \rho^2, S_j^1) > I & j=1, \dots, m_1 \\ R(\rho^1, \rho^2, S_{j'}^2) < I & j'=1, \dots, m_2 \end{cases}$$

Такое распознавание будем называть устойчивым распознаванием объектов обучения. Однако в ряде случаев условия (*) могут не выполняться, тем не менее может быть сформулирована приемлемая (с содержательной точки зрения) процедура распознавания. Такая ситуация имеет место, если выполняются условия

$$(**) \quad R(\rho^1, \rho^2, S_j^1) > R(\rho^1, \rho^2, S_{j'}^2) \text{ для всех}$$

$j=1, \dots, m_1, j'=1, \dots, m_2$. Положим

$$\epsilon_1 = \min_{j=1, \dots, m_1} R(\rho^1, \rho^2, S_j^1), \quad \epsilon_2 = \max_{j'=1, \dots, m_2} R(\rho^1, \rho^2, S_{j'}^2)$$

тогда (**) означает, что $\epsilon_1 > \epsilon_2$. Сформулируем решающее правило:

- а) при $R(\rho^1, \rho^2, S) \geq \epsilon_1$ S_n относится к первому классу;
 б) при $R(\rho^1, \rho^2, S) \leq \epsilon_2$ S_n относится ко второму классу;
 в) при $\epsilon_2 < R(\rho^1, \rho^2, S) < \epsilon_1$ S_n не распознается.

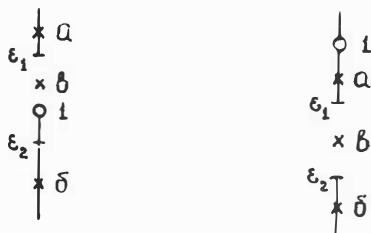


Рис. 3

На рис. 3 схематически показаны случаи (а), (б), (в).

Распознавание по этому правилу при невыполнении условий (*) называется неустойчивым. С содержательной точки зрения устойчивое распознавание представляется более приемлемым, чем неустойчивое, своей интуитивной очевидностью.

Наряду с таблицами T_1, T_2 рассмотрим их типизированные варианты $\tilde{T}_1 = (\tilde{t}_{ij})$ и $\tilde{T}_2 = (\tilde{t}_{ij}^x)$, где

$$\tilde{t}_{ij}^x = -|t_{ij}^k + t_{j-1}^k|, k=1, 2$$

Таким образом $\tilde{t}_{ij}^k = 1$, если $t_{ij}^k = \tilde{t}_{ij}^k$ и $\tilde{t}_{ij}^k = 0$ в противном случае.

Таблица \tilde{T}_K получается из T_K инверсией тех ее столбцов, для которых значение типичной строки равняется нулю. Сумма единиц i -ой строки таблицы \tilde{T}_K равняется числу совпадений i -ой строки T_K с типичной строкой таблицы T_K . Сумма единиц j -го столбца \tilde{T}_K показывает, на скольких из объектов $s_1^K, \dots, s_{m_K}^K$ признак x_j принимает свое типичное значение.

Пусть $\tilde{\pi}_1^K, \dots, \tilde{\pi}_n^K, \tilde{\omega}_1^K, \dots, \tilde{\omega}_{m_K}^K$ - "качельные" нагрузки столбцов и строк таблицы \tilde{T}_K .

Учитывая особенности кодирования признаков таблицы \tilde{T}_K , естественно интерпретировать величины $\tilde{\omega}_1^K, \dots, \tilde{\omega}_{m_K}^K$ как меры типичности объектов s_i^K в группе объектов $s_1^K, \dots, s_{m_K}^K$. Такое использование $\tilde{\omega}_i^K$ проводилось на ряде геологических примеров и дало результаты, хорошо согласующиеся с содержательной трактовкой типичности геологических объектов.

С содержательной точки зрения величина $\tilde{\pi}_j^K$ может быть истолкована как мера характеристичности признака x_j для группы объектов $s_1^K, \dots, s_{m_K}^K$, совмещающая в себе как количественную, так и качественную оценки выполнения типичного значения признака. Используем полученные нагрузки столбцов $\tilde{\pi}_j^K$ в описанной выше диагностической схеме, положив

$$P_I^K = \left(\frac{\tilde{\pi}_1^K}{\sum_{j=1}^n \tilde{\pi}_j^K}, \dots, \frac{\tilde{\pi}_n^K}{\sum_{j=1}^n \tilde{\pi}_j^K} \right) = (\tilde{\pi}_1^{K,H}, \dots, \tilde{\pi}_n^{K,H}), \quad Z_I^K = z(P_I^K, S), \quad R_I(s) = \frac{z(P_I^K, S)}{z(P_I^K, S)}$$

Указанную диагностическую схему реализует программа П2, которая вычисляет величины $R_I(s)$, $R_{II}(s)$, упомянутые выше, для испытуемых объектов и объектов экзамена S . Для контроля распознавания рекомендуется в число испытуемых объектов вводить все строки таблицы П1, П2, т.к. это позволяет убедиться в разделении классов. Кроме того, программа вычисляет величины R_{III} и R_{IV} ; аналогичные R_I и R_{II} соответственно, но учитывающие совпадения с репликами только по единицам.

Л и т е р а т у р а .

1. Ю.Л.Васильев, А.Н.Дмитриев. Спектральный подход к сравнению объектов, охарактеризованных набором признаков. ДАН СССР, 206, № 6, 1972.
2. Алгоритмический язык Алгол-60. Пересмотренное сообщение. Под ред. А.П.Ершова, М.Р.Шура-Бура. М., "Мир", 1965.

3. А.П.Ершов, Г.И.Кожухин, И.В.Поттосин. Руководство к пользованию системой Альфа. ВЦ СО АН СССР, Новосибирск, 1968.
4. Ляшенко В.Ф. Программирование для ЦВМ с системой команд типа М-220. М., "Советское радио", 1974.
5. Волков П.П., Дмитриев А.Н., Ногаева Т.М., Пантелеева В.И. Дифференциальная диагностика заболеваний головного мозга логико-дискретным методом. Сб. "Проблемы моделирования психической деятельности". вып.2, Новосибирск, 1968.

ПРОГРАММА ПИ "КАЧЕЛИ ДЛЯ БИНАРНЫХ ТАБЛИЦ"
Составлена на языке Альфа [3] для ЭВМ типа М-220, М-222.

Кандыба В.Н.

Назначение.

Программа реализует алгоритм метода согласованных оценок "Качели" [1], по которому вычисляются нагрузки строк (ω_i) и столбцов (π_j) бинарной таблицы t_{ij} . Таблица составлена из m объектов, характеризующихся n признаками.

Инструкция к пользованию.

Порядок постановки перфокарт.

1. Паспорт задачи.
2. Альфа-схема (I-34 п-к, 35 к Σ).
3. $m, k \Sigma$ (36 п-к).
4. $n, k \Sigma$ (37 п-к).
5. t_{ij} (38 ...п-к) с $K \Sigma$

где m - число строк, n - число столбцов. После каждого ввода прокладываются три пустых перфокарты. Информация кодируется следующим образом: 1. Каждая строка массива t разбивается по триадам и записывается в восьмеричных кодах, начиная с k .

2. Каждая строка $t[i, j]$ записывается с новой строки.

3. Массив перфорируется в режиме "команда".

4. В теле программы меняется выражение 4α , в котором переменной является величина n , определяющая длину одной полной строки таблицы. Выражение 4α имеет вид:

$$\{ \text{логический массив } t [1:m, 1:n] ; \quad ,$$

где вместо n ставится 45, 90, ... в зависимости от числа столбцов, т.к. каждая строка перфокарты имеет 45 разрядов. Число строк m и число столбцов n перфорируется в режиме "число" (см. контрольный пример).

В ы в о д осуществляется в следующем порядке: $\omega_i, \pi_j, \omega_i^*$, (упорядоченные по убыванию величины нагрузок), K - порядковые номера $\omega_i, \pi_j, \omega_i^*$, (упорядоченные по убыванию величины нагрузок), l - порядковые номера.

Время трансляции \approx равно пять минут, решение на М-220 при $m=9, n=9$ менее одной минуты. Объем информации не должен превышать 3000. При вычислении размера таблицы требуется учитывать условие,

чтобы $(m \times n): 45 < 3000$. Таблица $m=63, n=167$ с трансляцией занимает на БЭСМ три минуты.

Контрольный пример.

Имеется таблица, состоящая из 5 строк и 9 столбцов ($m=5, n=9$).

№	1	2	3	4	5	6	7	8	9
1	0	0	1	0	0	0	1	0	0
2	0	0	1	0	0	1	0	1	0
3	1	0	0	0	0	1	1	0	0
4	0	0	1	0	0	0	1	0	0
5	0	0	1	0	0	0	0	0	1

Закодированная информация имеет вид:

Сменное выражение 4α :

{ логический массив $t [I:m, I:45]$;
перфокарта (№ 36)

π	КОП	A1	A2	A3	
+	01	500	-	-	к Σ
+	01	500	-	-	

$m = 5$

перфокарта (№ 37)

π	КОП	A1	A2	A3	
+	01	900	-	-	к Σ
+	01	900	-	-	

$n = 9$

Массив (п/к № 38)

π	КОП	A1	A2	A3
1	04	-	-	-
1	12	-	-	-
4	14	-	-	-
1	04	-	-	-
1	01	-	-	-

t_{ij}

+ к Σ_t

Выдача на печать производится в следующем порядке:
 ω_i и π_j соответствуют объектам и признакам в порядке, представленном в таблице (контрольном примере)), а ω_i^* и π_j^* - упорядочены согласно K и l .

№ объектов	ω_i	ω_i^*	K
1	0,999	1,000	2
2	1,000	0,999	1
3	0,853	0,999	4
4	0,999	0,853	3
5	0,664	0,664	5

N	1	2	3	4	5	6	7	8	9
π_j	0,233	0,0	1,0	0,0	0,0	0,506	0,779	0,273	0,181
π_j^*	1,0	0,779	0,506	0,273	0,233	0,181	0	0	0
l	3	7	6	8	1	9	4	5	2

ТЕКСТ ПРОГРАММЫ "КАЧЕЛИ I"

1. начало
2. целые m, n, i, j, b ;
3. ввод (m, n) ;
4. ЛОГИЧЕСКИЙ массив $t [1 : m, 1 : n]$;
5. вещественный массив $x [1 : m], y [1 : n]$; ЦЕЛЫЙ массив $k [1 : m], l [1 : n]$;
6. вещественные $c, p, p1, a$;
7. ввод (t) ;
8. для $j := 1, \dots, n$ ЦИКЛ $y [j] := 1 ; p := 0 ;$
9. М $1 : \text{ для } i := 1, \dots, m$ ЦИКЛ $\{ x [i] := 0 ; \text{ для } j := 1, \dots, n$
10. ЦИКЛ если $t [i, j]$ то $x [i] := x [i] + y [j]$;
11. $c := x [1]$; для $i := 2, \dots, m$ ЦИКЛ
12. если $c < x [i]$ то $c := x [i]$;

13. для $i := I, \dots, m$ цикл $x[i] := x[i] / c$;
14. для $j := I, \dots, n$ цикл { у $y[j] := 0$; для $i := I, \dots, m$ цикл
15. если $t[i, j]$ то $y[j] := y[j] + x[i]$ } ;
16. $c := y[I]$; для $j := I, \dots, n$ цикл
17. если $c < y[j]$ то $c := y[j]$; $pI := p$; $p := c$;
18. для $j := I, \dots, n$ цикл $y[j] := y[j] / c$;
19. если $abs(p - pI) > 0.00001$ то на МТ ;
20. вывод (x, y) ;
21. для $i := I, \dots, m$ цикл
22. $k[i] := i$;
23. для $i := I, \dots, m - I$ цикл
24. для $j := i + I, \dots, m$ цикл
25. если $x[i] < x[j]$ то { $a := x[i]$; $x[i] := x[j]$; $x[j] := a$;
26. $b := k[i]$; $k[i] := k[j]$; $k[j] := b$ } ;
27. для $j := I, \dots, n$ цикл
28. $l[j] := j$;
29. для $j := I, \dots, n - I$ цикл

30. для $i := j + 1, \dots, n$ цикл
 31. если $y[j] < y[i]$ то { $a := y[j]; y[j] := y[i]; y[i] := a;$
 32. $b := 1[j]; 1[j] := 1[i]; 1[i] := b;$
 33. вывод ($x, k, y, 1$) }
 34. конец *

19

Таблица распределения памяти

	Начало	Конец	Длина
Программа	0063	0743	0661
Числовые константы	0744	0770	0025
формируемые константы	0771	1051	0061
Сколяры	1052	1110	0037
Рабочие ячейки	1111	1116	0006
Массивы не распределены.			
Первое рабочее поле	1117	7306	6170
Рабочее поле ИС	7307	7500	0171
Адреса программных остановов	0743		

ПРОГРАММА П2 "РАСЧЕТ КОЭФФИЦИЕНТОВ"

составлена на языке Алгол для ЭВМ типа М-220, М-222

Кавдыба В.Н.

Назначение. Программа П2 вычисляет "качельные" нагрузки столов $\tilde{\omega}_j$ и строк $\tilde{\omega}_i$; "типизированных" вариантов Т1 и Т2 таблиц Т1 и Т2, а также предназначенные для распознавания коэффициенты $R_I, R_{II}, R_{III}, R_{IV}$ для каждого испытуемого объекта (строки Т3). Программа позволяет решать одновременно несколько задач.

Инструкция к пользованию.

Порядок постановки перфокарт.

- | | |
|--|--|
| 1. Альфа-схема (I+26 α) K Σ | |
| 2. C1, K Σ - число задач | 9. T3, K Σ - массив проб |
| 3. n, K Σ - число столов | 10. T1, K Σ - массив эталонов I класса |
| 4. m1, K Σ - число строк в Т1 | 11. m1, K Σ - число строк в Т1 |
| 5. m2, K Σ - число строк в Т2 | 12. T2, K Σ - массив эталонов II класса |
| 6. m3, K Σ - число строк в Т3 | 13. m2, K Σ - число строк в Т2 |
| 7. { ϕ_1, ϕ_2 , K Σ - реплики Т1 и Т2 | |
| 8. { ϕ_1, ϕ_2 , K Σ - реплики Т1 и Т2 | |

Информация кодируется следующим образом:

1. Каждая строка массива $t[i, j]$ разбивается по триадам и записывается в восьмеричных кодах, начиная с κ .

2. Каждая строка $t[i, j]$ записывается с новой строки.

3. Массивы перфорируются в режиме "команда".

4. В теле программы меняется выражение -3α , которое имеет вид: { логический массив $t[I:m, I:n^*]$, $t3[I:m3, I:n^*]$,

Вместо n^* указывается величина $n^* = z \cdot 45$, где z - количество строк перфокарты, занятых под одну строку таблицы. Например, для Т1 контрольного примера $n^* = I \times 45$.

Недостающие разряды заполняются нулями. Реплики ϕ_1 и ϕ_2 рассматриваются как одна строка, сначала ϕ_1 , затем ϕ_2 , которая разбивается по триадам и кодируется в режиме "команда". Они вводятся дважды с суммами. Число задач (C1), число строк (m) и число столов (n) перфорируются в режиме "число" (см. контрольный пример).

Каждый ввод прокладывается тремя пустыми перфокартами.

Вывод информации производится в следующем порядке:

для задачи № I:

ω_i^1 - качельные нагрузки строк Т1

x) по ϕ_1 Т1 и Т2 "типизируются" для расчета $\tilde{\omega}_i, \tilde{\omega}_j$ по ϕ_2 вычисляются $R_I, R_{II}, R_{III}, R_{IV}$.

$\tilde{\pi}_j^1$ - качельные нагрузки столбцов T1

$\tilde{\omega}_i^2$ - качельные нагрузки строк T2

$\tilde{\pi}_j^2$ - качельные нагрузки столбцов T2

$R_{I_1}, R_{I_2}, R_{II}, R_{IV}$ - по м3 чисел каждый.

Время трансляции - 8 минут. Решение при T1 = (5 x 9), T2 = (9 x 9), T3 = (19 x 9) занимает менее одной минуты. Предварительный расчет объема информации можно производить по формуле, учитывающей размеры всех динамических массивов:

$$n [m^{(x)} + m3 + 2] : 45 + m^{(x)} + m1 + m2 + 5n + 4 \cdot m3 = 2644.$$

ж) если $m1 < m2$, то $m=m2$, иначе $m=m1$.

При решении нескольких задач одновременно порядок постановки перфокарт сохраняется. При этом данные второй задачи вводятся за перфокартой m2, KΣ для I-й задачи в обычном порядке, начиная с n, KΣ. Порядок выдачи на печать результатов решения 2-й, ..., n-й задачи дублирует порядок выдачи по первой задаче:

для задачи № 2

$\tilde{\omega}_i^1, \tilde{\pi}_j^1, \tilde{\omega}_i^2, \tilde{\pi}_j^2, R_{I_1}, R_{I_2}, R_{II}, R_{IV}$, и т.д.

Контрольный пример

В контрольном примере представлено решение одной задачи (сI=I). Имеется 3 таблицы: T1 и T2 - таблицы эталонов и T3 - таблица проб.

T1							T2							T3						
№	I	2	3	4	5	6	№	I	2	3	4	5	6	№	I	2	3	4	5	6
I	0	0	I	I	I	I	I	I	0	I	I	I	I	I	0	0	0	0	I	I
2	0	I	0	0	0	I	2	I	0	0	0	0	I	2	0	I	I	0	0	I
3	I	I	I	0	0	I	3	0	0	I	0	0	I	3	I	I	0	0	0	I
ΦI	0	I	I	0	0	I	Φ2	I	0	I	0	0	I	4	0	I	0	I	0	I

Закодированная информация имеет вид:

сменное выражение 3α : {логический массив т [I:m, I:45], т3 [I:m3, I:45]}

К о д и р о в к а н а б л а н к а х

π	КОП	A1	A2	A3
+	0I	I00	-	-
+	0I	I00	-	- KΣ

перфокарта сI, KΣ

π	КОП	A1	A2	A3
+	01	600	-	-
+	01	600	-	- K Σ

перфокарта н,к Σ

π	КОП	A1	A2	A3
+	01	300	-	-
+	01	300	-	- K Σ

перфокарта м1,к Σ

π	КОП	A1	A2	A3
+	01	300	-	-
+	01	300	-	- K Σ

перфокарта м2, к Σ

π	КОП	A1	A2	A3
+	01	400	-	-
+	01	400	-	- K Σ

перфокарта м3, к Σ

π	КОП	A1	A2	A3
1	70	-	-	-
2	10	-	-	-
7	10	-	-	-

массив T1

+ K Σ_{T1}

π	КОП	A1	A2	A3
5	70	-	-	-
4	10	-	-	-
1	10	-	-	-

массив T2

+ K Σ_{T2}

π	КОП	A1	A2	A3
0	30	-	-	-
3	10	-	-	-
6	10	-	-	-
2	50	-	-	-

массив T3

+ K Σ_{T3}

π	КОП	A1	A2	A3
3	15	1000	0000	0000

($\Phi 1 + \Phi 2$) = Φ (реплика)

Результаты счета представлены в таблицах:

$\tilde{\omega}_i^1$ и $\tilde{\pi}_j^1$ для таблицы T1

№	1	2	3	№	1	2	3	4	5	6
$\tilde{\omega}_i^1$	0,219	0,390	0,390	$\tilde{\pi}_j^1$	0,133	0,171	0,133	0,171	0,171	0,219

$\tilde{\omega}_i^2$ и $\tilde{\pi}_j^2$ для таблицы T2

№	1	2	3	№	1	2	3	4	5	6
$\tilde{\omega}_i^2$	0,288	0,356	0,356	$\tilde{\pi}_j^2$	0,136	0,212	0,136	0,151	0,151	0,212

Коэффициенты $R_I, R_{\bar{I}}, R_{\bar{II}}, R_{\bar{IV}}$

	1	2	3	4
R_I	0,910	1,536	1,125	1,913
$R_{\bar{I}}$	1,000	1,500	1,000	2,000
$R_{\bar{II}}$	1,033	1,502	1,119	1,840
$R_{\bar{IV}}$	1,000	1,500	1,000	2,000

ТЕКСТ ПРОГРАММЫ П2

начало целый м, н, и, ж, к, м 1, м2, м3, с, с 1; ввод (с 1); начало для с:= 1, ..., с1
цикл начало ввод (н, м 1, м2, м3); если м2 > м 1 то м:=м2 иначе м:=м 1;
3а: начало логический массив т[1 : м, 1:45] , т3 [1:м3, 1:45] , ф 1, ф [1:2, 1:н] ;
вещественный массив а[1:м] , а 1[1:м 1], а2 [1:м2] , у[1 :н], у1 [1:2, 1:н] , р, р 1[1:н] ,
 ч 1, ч2, в1 , в2 [1:м3] ; вещественный в ; ввод (ф 1, ф, т3); для к:=1 , 2 цикл
начало ввод (т, м) ; для и:= 1 , ..., м цикл для ж:= 1 , ..., н цикл если
 т [и, ж] and ф 1 [к, ж] or not т[и, ж] and not ф 1 [к, ж] то т [и, ж] := истина иначе
 т [и, ж] := ложь ; для ж := 1 , ..., н цикл начало у [ж] := 1 ; р [ж] := 0 конец ;
 Мс1: для и:= 1 , ..., м цикл начало а [и] := 0 ; для ж:= 1, ..., н цикл если
 т [и, ж] то а [и] := а [и] + у [ж] конец ; в := 0 ; для и:= 1 , ..., м цикл
 в := в + а [и] ; для и:= 1 , ..., м цикл а [и] := а [и] / в ; для ж := 1, ..., н
цикл начало у [ж] := 0 ; для и:= 1 , ..., м цикл если т [и, ж] то у [ж] := у [ж] + а [и]
конец ; в := 0 ; для ж:= 1, ..., н цикл в := в + у [ж] ; для ж:= 1, ..., н цикл начало
 у [ж] := у [ж] / в , р1[ж] := р [ж] ; р [ж] := у [ж] конец; для ж:= 1 , ..., н цикл если

$\text{abs} (p [j] - p_1 [j]) > 0,00001$ то на $M \epsilon 1$; если $k=1$ то начало для $j:=1, \dots, n$
цикл $y_1 [1, j] := y [j]$; для $i:=1, \dots, m_1$ цикл $a_1 [i] := a [i]$; вывод (a_1, y)
конец иначе начало для $j:=1, \dots, n$ цикл $y_1 [2, j] := y [j]$; для $i:=1, \dots, m_2$
цикл $a_2 [i] := a [i]$; вывод (a_2, y) конец конец ; для $k:=1, 2$ цикл начало для
 $i:=1, \dots, m_3$ цикл начало $v_1 [i] := v_2 [i] := 0$; для $j:=1, \dots, n$ цикл если $(\phi [k, j]$
 $\text{and } t_3 [i, j]) \text{ or not } \phi [k, j] \text{ and not } t_3 [i, j]$ то начало $v_1 [i] := v_1 [i] + y_1 [k, j]$;
 $v_2 [i] := v_2 [i] + 1$ конец конец ; для $i := 1, \dots, m_3$ цикл если $k=1$ то начало
 $ch_1 [i] := v_1 [i]$; $ch_2 [i] := v_2 [i]$ конец иначе начало $ch_1 [i] := ch_1 [i] / v_1 [i]$; $ch_2 [i] := ch_2 [i]$
 $/ v_2 [i]$ конец конец ; вывод (ch_1, ch_2) ; для $k:=1, 2$
цикл начало для $i:=1, \dots, m_3$ цикл начало $v_1 [i] := v_2 [i] := 0$; для $j:=1, \dots, n$
цикл если $\phi [k, j] \text{ and } t_3 [i, j]$ то начало $v_1 [i] := v_1 [i] + y_1 [k, j]$; $v_2 [i] := v_2 [i] + 1$
конец конец ; для $i:=1, \dots, m_3$ цикл если $k=1$ то начало $ch_1 [i] := v_1 [i]$;
 $ch_2 [i] := v_2 [i]$ конец иначе начало если $(v_1 [i] + v_2 [i]) < 10 - 10$ то $ch_1 [i] := ch_2 [i] := 0$
иначе начало $ch_1 [i] := ch_1 [i] / v_1 [i]$; $ch_2 [i] := ch_2 [i] / v_2 [i]$ конец конец конец ; вывод
 (ch_1, ch_2) конец конец конец конец *

Таблица распределения памяти

	Начало	Конец	Длина
Программа	0063	2170	2106
Числовые константы	2171	2217	0027
Формируемые константы	2220	2326	0107
Скаляры	2317	2434	0106
Рабочие ячейки	2435	2444	0010
Массивы не распределены			
Первое рабочее поле	2445	7306	4642
Рабочее поле ИС	7307	7500	0171
Адреса программных остановов 2170			

ПРОГРАММА ПЗ " СИММЕТРИЯ "

Составлена в машинных кодах для ЭВМ типа М-220, М-222.

Смертин Е.А.

Назначение.

Программа ПЗ предназначена для вычисления качельных нагрузок строк и столбцов [I] для таблиц, составленных из (0 и I), а также для таблиц, составленных из (I и -I).

Инструкция к пользованию.

I. Подготовка исходных данных. Исходная таблица T, для которой вычисляется нагрузки строк и столбцов, должна иметь размерность не больше 45 x 180, т.е. должно выполняться следующее условие

$$\min(m,n) \leq 45$$

$$\max(m,n) \leq 180$$

где m - число строк, а n - число столбцов в T.

Пусть таблица S равна

$$T, \text{ если } n \leq 45$$

$$T', \text{ если } n > 45,$$

причем число строк S обозначим p, а число столбцов - z; тогда, очевидно, $p \leq 180$, а $z \leq 45$.

Вычисление нагрузок для T проводится с помощью таблицы S, которая кодируется на перфокартах, начиная с п/к 7 и вводится в МОЗУ, начиная с ячейки 0322. Итак, исходная информация представляется в виде п/к 7: $k \leq 22$, первоначально заполненных нулями.

п/к 7	π	КОП	A1	A2	A3	
	0	0 0	0322	0000	0000	КА
	S_1^1	S_2^1	S_3^1	S_{45}^1

	S_1^{11}	S_2^{11}	S_3^{11}	S_{45}^{11}

п/к 8	π	КОП	A1	A2	A3	
	S_1^{12}	S_2^{12}	S_3^{12}	S_{45}^{12}
	S_1^{23}	S_2^{23}	S_3^{23}	S_{45}^{23}

п/к 22	
	π	КОП	A1	A2	A3	
	S_1^{180}	S_2^{180}	S_3^{180}	S_{45}^{180}

где S_i^j есть $\begin{cases} 0, \text{если } i > z, \\ 0, \text{если } j > p, \\ \text{элемент, стоящий в } i\text{-м столбце } j\text{-й строки матрицы } S \\ \text{если } i \leq z \text{ и } j \leq p \end{cases}$

Кроме п/к 7-22 составляется п/к 35 вида:

п/к 35	КОП	AI	A2	A3	
0	0 0	03I0	0000	0000	KA
0	0 0	q-I	0000	0000	
0	0 0	007I	0000	0000	KA
I	I 2	p-I	0064	000I	
0	0 0	0074	0000	0000	KA
I	I 2	z-I	0062	000I	
0	0 0	0I07	0000	0000	KA
I	I 2	p-I	0I02	000I	
0	0 0	0I13	0000	0000	KA
I	I 2	z-I	0I00	000I	
		б/м			

Здесь q - число циклов вычисления нагрузок, p - число строк, а z - столбцов таблицы S . q - число циклов обычно берется равным: для (0,1)-ых таблиц - 10_{10} или 12_8
для (1,-1)-ых таблиц - 45_{10} или 55_8

В том случае, когда таблица T состоит из (1,-1), то она предварительно перекодируется в (0,1)-ую путем замены (-1) на 0. Обработка новой таблицы проводится с помощью той же самой программы, но из нее удаляется п/к 34^a и добавляется п/к 36, имеющая вид:

№ яч.	л	КОП	AI	A2	A3	
			0050			KA
0050			03I2		03I7	
005I	56			00II	0050	
			00II			KA
00II	55		03I7	00I0		
00I2	36			00I5		
00I3	I6		00I4	750I	76I0	
00I4	52		I437	0027	I777	
00I5	74		0057	03I7	03I7	
00I6	56			0052		
			б/м			

Кроме того, в п/к 35 добавляются 2 следующие строки:

π	КОП	AI	A2	A3	
-	--	0010	-	-	КА
-	--	-	-	0017	

На печать выдаются q циклов последовательных приближений нагрузок столбцов и строк.

После теста печати выдаются нагрузки столбцов π_j (z значений), нули ($45-z$) и нагрузки строк ω_i (p значений) и нули ($180-p$).

Контрольный пример

Имеется таблица, состоящая из 5 строк и 9 столбцов.

№	1	2	3	4	5	6	7	8	9	ω_i
1	0	0	1	0	0	0	1	0	0	0,221
2	0	0	1	0	0	1	0	1	0	0,221
3	1	0	0	0	0	1	1	0	0	0,189
4	0	0	1	0	0	0	1	0	0	0,221
5	0	0	1	0	0	0	0	0	1	0,147

} результаты
счета

π_j									
0,079	0	0,336	0	0	0,170	0,262	0,092	0,061	

} результаты счета
Время счета - 2мин.

Закодированная информация имеет вид:

п/к7	π	КОП	AI	A2	A3	п/к 35	π	КОП	AI	A2	A3
	0	00	0322	-	-	КА	0	00	0310		КА
	1	04							0011		
	1	12					0	00	0071		КА
	4	14					1	12	0004	0064	0001
	1	04					0	00	0074		КА
	1	01					1	12	0010	0062	0001
	0						0	00	0107		КА
	0						1	12	0004	0102	0001
	0						0	00	0113		КА
	0						1	12	0010	0100	0001
	0								б/м		
	0										

ТЕКСТ ПРОГРАММЫ ПЗ

№	п/к														
1	0	00	000I	0000	0000	КА	5	0	00	0000	0000	0000			
	0	52	0000	0000	0026			0	74	0057	032I	032I			
	I	00	0000	0000	I437			I	12	0054	0062	000I			
								0	00	0000	0000	0000			
	I	I2	0340	0002	000I			0	00	0000	0000	0000			
	0	56	0000	0020	0000			4	00	I437	0000	0320			
	0	00	0020	0000	0000	КА		4	52	0000	0000	0I10			
	0	50	00I3	0000	7767			0	00	03I2	0000	032I			
	0	70	7500	0020	0000			2	55	032I	0322	0000			
	0	I6	0023	750I	76I0			0	36	0000	0I06	0000			
	0	52	I076	0042	I436			2	0I	0320	II53	0320			
	0	I6	0025	0060	0075			0	56	0000	0I07	0000			
	0	I6	0026	0076	0II4										
	2	0	00	0000	0000	0000			6	2	02	0320	II53	0320	
		0	00	0000	0000	032I				I	I2	0263	0I02	000I	
		5	00	I437	0000	I076				0	00	0000	0000	0000	
		2	0I	032I	I437	032I				0	74	0057	032I	032I	
I		I2	0054	0030	000I		I	00		0320	0000	I437			
0		52	0000	0000	0037		I	I2		0054	0077	000I			
5		04	I076	032I	I076		0	00		03I0	0000	0000	КА		
5		00	I076	0000	I437		0	00		0005	0000	0000			
I		I2	0054	0034	000I		0	00		000I	0000	0000			
0		00	0000	0000	0000		4	00		0000	0000	0000			
0		00	0000	0000	032I										
5		00	I5I4	0000	II53										
3		2	0I	032I	I5I4	032I		8		0	00	0000	0000	0000	
	I	I2	0263	004I	000I		0		00	0000	0000	0000			
	0	52	0000	0000	0052		0		00	0000	0000	0000			
	5	04	II53	032I	II53		0		00	0000	0000	0000			
	5	00	II53	0000	I5I4		0		00	0000	0000	0000			
	I	I2	0263	0045	000I		0		00	0000	0000	0000			
	0	I6	005I	750I	76I0		0		00	0000	0000	0000			
	0	52	I437	0027	I777		0		00	0000	0000	0000			
	0	00	0000	0000	0000		0		00	0000	0000	0000			
	I	00	0000	0000	I437		0		00	0000	0000	0000			
	I	I2	0340	0053	000I		9		0	00	0000	0000	0000		
	0	33	03I0	03II	03I0				0	00	0000	0000	0000		
									0	00	0000	0000	0000		
4	0	76	0000	0024	0000			0	00	0000	0000	0000			
	0	77	0000	0000	0000			0	00	0000	0000	0000			
	0	00	03I2	0000	032I			0	00	0000	0000	0000			
	0	52	0000	0000	0076			0	00	0000	0000	0000			
	4	00	I076	0000	0320			0	00	0000	0000	0000			
	4	52	0000	0000	0072			0	00	0000	0000	0000			
	2	55	032I	0322	0000			0	00	0000	0000	0000			
	0	36	0000	0070	0000			0	00	0000	0000	0000			
	5	0I	I5I4	0320	I5I4			0	00	0000	0000	0000			
	0	56	0000	007I	0000			0	00	0000	0000	0000			
	5	02	I5I4	0320	I5I4										
	I	I2	0263	0064	000I										

п-к 7 - информация

№ п/к
26

0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000

26^a

0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000

26^б

0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000

27

0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000

27^a

0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000

27^б

0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000

28

0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000
0	00	0400	0000	0000

29

0	00	003I	0000	0000	кА
4	56	I437	0	II5	0320
0	00	0042	0000	0000	кА
4	56	I5I4	0I20	0320	
0	00	0II5	0000	0000	кА
0	03	0320	0000	0320	
0	0I	032I	0320	032I	
0	56	0000	0032	0000	
0	03	0320	0000	0320	
0	0I	032I	0320	032I	
0	56	0000	0043	0000	
0	00	0000	0000	0000	

№ П/К

30 0 00 0071 0000 0000 KA
 I I2 0020 0064 0001
 0 00 0074 0000 0000 KA
 I I2 0024 0062 0001
 0 00 0107 0000 0000 KA
 I I2 0020 0102 0001
 0 00 0113 0000 0000 KA
 I I2 0024 0077 0001

34 0 00 0022 0000 0000 KJ
 0 I6 0023 7501 7610
 0 00 0027 0000 0000 KJ
 0 00 0000 0000 0321
 0 00 0050 0000 0000 KJ
 0 I6 0051 7501 7610
 0 00 0155 0000 0000 KJ
 0 00 0131 0000 0310
 0 I6 0157 0024 0057
 0 77 0000 0000 0000

31 0 00 0077 0000 0000 KA
 0 00 0312 0000 0321
 4 00 I437 0000 0320
 4 52 0000 0000 0110
 0 00 0113 0000 0000 KA
 I I2 0024 0100 0001

34^B 0 00 0070 0000 0000 K
 0 00 0000 0000 0000
 0 00 0106 0000 0000 K
 0 00 0000 0000 0000

32 0 00 0131 0000 0000 KA
 0 00 0007 0000 0000
 0 00 0001 0000 0000
 0 I6 0051 7501 7610
 0 52 I437 0027 I777
 0 00 0000 0000 0050
 0 00 0000 0000 0051
 0 00 0131 0000 0310
 0 I6 0141 0020 0057
 0 52 0000 0000 0000
 0 00 0133 0000 0050
 0 00 0134 0000 0051

K Σ 2 36 7006 7173 2643

33 0 00 0132 0000 0310
 0 I6 0146 0024 0057
 0 00 0000 0000 0050
 0 00 0000 0000 0051
 0 00 0131 0000 0310
 0 I6 0152 0024 0057
 I I2 0007 0142 0001
 0 00 0133 0000 0050
 0 00 0134 0000 0051
 0 I6 0156 0024 0057
 0 77 0000 0000 0000
 0000 0004 0000 0000 KA

ПРОГРАММЫ "ТЕСТОВЫХ ОЦЕНОК" (ТЕСТЫ, ТЕСТОРЫ)

Серия программ П1, П2, П3, П4, П5, П6 предназначена для изучения группы предметов или явлений, характеризуемых набором признаков, принимающих значения I, 0, "-" так, что m объектов с n признаками полностью описываются таблицей

$$T = \{t_{ij}\}, \quad t_{ij} \in \{0, I, -\}, \quad \begin{matrix} j = 1, 2, \dots, m \\ i = 1, 2, \dots, n \end{matrix}$$

В основу программ положен метод изучения объектов с помощью тестов и тесторов, предложенный и описанный в [1, 2, 5].

В качестве меры важности отдельного признака принимается его информационный вес $P_{(i)}$, определяемый так: $P_{(i)} = \frac{K_i}{K}$, где K - общее число тупиковых тестов таблицы T , K_i - число тупиковых тестов, в которые входит i -ый столбец (признак).

Таким образом, для вычисления информационных весов нужно перебрать все тупиковые тесты таблицы T , что является достаточно трудоемким процессом, предъявляющим повышенные требования к ЭВМ (число операций $\sim c_1 m^2 \cdot 2^n + c_2 k^3$, где c_1, c_2 - константы).

В связи с этим, вся серия программ, за исключением П1, составлена для машины БЭСМ-6.

Во всех программах П1-П6 для нахождения тупиковых тестов использован алгоритм, существенно уменьшающий полный перебор [4]. Однако, эффективность его зависит от конфигурации таблицы T , от количества тупиковых тестов и их длин. Это обстоятельство не позволяет точно оценить его трудоемкость.

Эксплуатация программы П1 на машине М-220 показала, что таблицы с $m \approx 15+20$, $n \approx 20+25$, обрабатываются за 30-40 минут.

Программа П2 на машине БЭСМ-6 просчитывает таблицу с $m=15+20$, $n \approx 30+35$ примерно за 1 час.

Алгоритмы для П1-П6 были разработаны в ИМ СО АН СССР под руко-
водством д.ф.-и.н. Ю.И. Журавлева в 1967-1969 г.г.

§ 1. Описание алгоритма вычисления информационных весов (программы П1 и П2)

1. По исходной таблице $T = \{t_{ij}\}$ - составляется таблица T_I , строки которой суть результаты попарных сравнений всех строк из T .
2. Таблица T_I "уплотняется" за счет выбрасывания строк, для которых хотя бы одно сужение входит в T_I .
3. Так как порядок строк и столбцов в таблице T_I произволен, то для более эффективной (в смысле затрат машинного времени) работы алгоритма T_I переупорядочивается так, чтобы нули концентрировались в правом верхнем углу. Это достигается следующим образом. Отскивается строка S_1^* , содержащая минимальное число единиц. Эта строка выносится на первое место новой таблицы T^* , а z_1 столбцов таблицы T_I , которым соответствуют единицы в строке S_1^* , выносятся на первые z_1 мест.

Среди оставшихся строк ищется строка S_2^* , содержащая минимум единиц в оставшихся $(n - z_1)$ столбцах. Эта строка становится второй в T^* , и ее z_2 единиц в оставшихся $(n - z_1)$ рядах определяют следующие z_2 столбцов в таблице T^* и т.д., пока не будут исчерпаны все строки и столбцы.

Построенная таким образом T^* является исходной для выделения всех тупиковых тестов заданной таблицы T .

4. Пусть $(i_1, i_2, \dots, i_\ell)$ - тест. Набор $X = \{x_1, x_2, \dots, x_n\}$, у которого координаты с номерами i_1, i_2, \dots, i_ℓ равны единице, а все остальные - нулю, назовем тестовым набором.

Очевидно, что для каждого теста таблицы T из n столбцов есть соответствующий тестовый набор в множестве $M = \{X\}$ всевозможных наборов

$$X = \{x_1, x_2, \dots, x_n\}, \quad x_i \in \{0, 1\}, \quad i = 1, 2, \dots, n$$

Таким образом, задачу отыскания всех тупиковых тестов можно решить, перебрав все наборы $X \in M$ и выбрав из них все те и только те тестовые наборы, которые отвечают тупиковым тестам.

При этом оказывается, что часть наборов $X \in M$ выпадает из рассмотрения автоматически на том основании, что всякое расширение тестового набора не может соответствовать тупиковому тесту, а всякое сужение нетестового набора и подавно является нетестовым.

Наборы просматриваются, начиная с $X = \{0, 0, \dots, 0\}$ в

порядке возрастания, если их рассматривать как двоичные изображения чисел от 0 до $(2^n - 1)$.

Обработка очередного набора $X = \{x_i\}$ распадается на два этапа.

I этап. Выясняется, является ли набор $X = \{x_i\}$ тестовым. Для этого в таблице T^k ищется строка s^* такая, чтобы $s^* \leq \bar{X}$ (поскольку таблица T^k упорядочена нужным образом, то нас интересует только первая такая строка).

Если такая строка $s^* = \{t_1^*, t_2^*, \dots, t_n^*\}$ существует, то очевидно, $X = \{x_i\}$ не является тестовым набором, причем если $t_i^* = 1$, $t_{i+1}^* = t_{i+2}^* = \dots = t_n^* = 0$, то все следующие за $X = \{x_i\}$ в порядке возрастания $(2^{n-i} - 1)$ наборов также нетестовые.

Если при просмотре таблицы T^k окажется, что ни для какой из строк соотношений $s^* \leq \bar{X}$ не имеет места, значит, X - тестовый набор. При этом, если $x_{2+1} = x_{2+2} = \dots = x_n = 0$, то очевидно, что тестовыми наборами являются и следующие $(2^{n-2} - 1)$ наборов. Однако, они также выпадают из рассмотрения, так как заведомо не соответствуют тупиковым тестам.

II этап. Проверка тестовых наборов на тупиковатость. Все тестовые наборы, соответствующие тупиковым тестам, запоминаются в оперативной памяти машины.

Каждый раз при получении очередного тестового набора $X = \{x_i\}$ просматривается таблица θ накопленных тупиковых тестов.

Если X не является расширением ни одного из наборов таблицы θ , то $X = \{x_i\}$ - новый тупиковый тест и таблица θ пополняется, т.к. ни один из получаемых в дальнейшем тестовых наборов не может быть сужением X (это утверждение следует из просмотра наборов X по возрастанию).

Таблица θ , полученная в результате просмотра всех наборов до $X = \{1, 1, \dots, 1\}$, содержит все тупиковые тесты исходной таблицы T и только их.

В результате работы III вычисляются и печатаются следующие величины:

а) $P_i = \frac{K_i}{K}$ - информационные веса;

б) $P'_i = \frac{A_i}{m} \cdot P_i$, где A_i - число единиц в i -ом столбце, $i = 1, 2, \dots, n$;

в) $J(s_j) = \sum_{i=1}^n P_i \cdot t_{ij}$; $J'(s_j) = \sum_{i=1}^n P'_i \cdot t_{ij}$, $j = 1, 2, \dots, m$; $t_{ij} \in \{0, 1\}$.

Если t_{ij} прочерк, то вместо "-" берется $\frac{1}{2}$.

Программа П1 позволяет также вычислять введенные в работе [3] отождествляющие веса признаков Q_i (Q - тестовые меры признаков). Согласно работе [3], если задана таблица T , то тупиковым Q -тестом для нее называется набор столбцов (i_1, i_2, \dots, i_k) такой, что в подтаблице, составленной из этих столбцов, любые две строки совпадают хотя бы по одному столбцу и никакой поднабор, входящий в (i_1, i_2, \dots, i_k) не обладает этим свойством.

Отождествляющий вес признака Q_i принимается равным K_i^a / K^a , где K_i^a - общее число тупиковых Q -тестов, в которые входит i -й столбец, а K^a - число всех тупиковых Q -тестов.

Для вычисления весов Q_i в программе П1 вместо таблицы T_I используется таблица T'_I , полученная заменой в T_I всех единиц нулями, а всех нулей - единицами.

Алгоритм нахождения всех тупиковых тестов в программе П2 в основном сохранен таким же, как и в П1, за исключением той его части, где проверяется тупиковость очередного теста (П этап).

Согласно определению тупикового теста, всякий поднабор его не является тестом, поэтому достаточно проверить все сужения исследуемого теста, поочередно опуская в нем по единице до тех пор, пока не встретится поднабор, являющийся тестом. Если такового не оказалось, то рассматриваемый тестовый набор соответствует тупиковому тесту.

Такая модификация алгоритма вызвана тем, что в задачах большого объема, решаемых с помощью П2, число тупиковых тестов достигает 50-70 тысяч. Хранить в памяти машины (или на внешних накопителях) такое количество информации не представляется возможным и целесообразным, т.к. многократный просмотр ее резко увеличивает затраты машинного времени.

§ 2. Вычисление тесторной информационной меры признаков (программа П3)

Согласно работе [1] тупиковым тестором для двух или нескольких таблиц T_1, T_2, \dots, T_j называется набор (i_1, i_2, \dots, i_k) столбцов такой, что никакие две подтаблицы среди подтаблиц T'_1, T'_2, \dots, T'_j , составленных только из этих столбцов, не имеют одинаковых строк и никакой поднабор (i'_1, i'_2, \dots, i'_k), входящий в (i_1, i_2, \dots, i_k)

не обладает этим свойством.

Тесторный вес признака $P^*(i)$ принимается равным отношению K_i^*/K^* , где K^* - общее число тупиковых тесторов, K_i^* - число тупиковых тесторов, в которые входит i -ый столбец.

Алгоритм для нахождения всех тупиковых тесторов аналогичен алгоритму, описанному в § I, и отличается от него только в первой части, когда производится попарное сравнение строк.

Очевидно, для нахождения тесторов не следует включать в таблицу T_I (см. § I) результаты сравнений строк каждой таблицы между собой.

С этой целью к программе П2 подключается дополнительный блок, который организует попарное сравнение всех строк из разных таблиц и только их. Программа позволяет также вычислять введенные в работе [3] отождествляющие веса признаков Q_i (Q - тестовые меры признаков). Согласно работе [3], если задана таблица T , то тупиковым Q -тестом для нее называется набор столбцов (i_1, i_2, \dots, i_r) такой, что в подтаблице, составленной из этих столбцов, любые 2 строки совпадают хотя бы по одному столбцу, и никакой поднабор, входящий в (i_1, i_2, \dots, i_r) не обладает этим свойством. Отождествляющий вес признака Q_i принимается равным K_i^a/K^a , где K^a - общее число тупиковых Q -тестов таблицы T , а K_i^a - число тупиковых Q -тестов, в которые входит i -ый столбец.

В соответствии с работой [6] тупиковым пакетом таблицы T называется набор столбцов (i_1, i_2, \dots, i_r) такой, что в подтаблице, составленной из этих столбцов, в каждой строке найдется хотя бы одна 1, и никакой поднабор, входящий в (i_1, i_2, \dots, i_r) не обладает этим свойством. Обозначим через \bar{T} таблицу, полученную из T заменой всех ее столбцов на зеркальные. Пакеты \bar{T} называются "0-пакетами" T . Пакетный вес признака A_i принимается равным отношению $\frac{F_i(T)}{F(T)}$, где $F(T)$ - общее число тупиковых пакетов таблицы T , а $F_i(T)$ - число тупиковых пакетов, в которые входит i -ый столбец. \bar{A}_i - аналогичная величина для "0-пакетов". - "Пакеты" иногда называют "I-пакетами".

§ 3. Диагностика объектов голосованием по тестам и тесторам (программы П4, П5)

Пусть, как и в предыдущих задачах, дана таблица $T = \{t_{ij}\}$, описывающая m эталонов с помощью n признаков. Пусть кроме этого задан

некоторый объект $S_j = \{t_{j_1}, t_{j_2}, \dots, t_{j_n}\}$, $t_{j_i} \in \{0, 1, -\}$, который надо отнести к одному из m классов, представителями которых являются выбранные эталоны [2]. Изложим алгоритм голосования, предложенный в [7, 8].

Пусть $\{i_1, i_2, \dots, i_\ell\}$ - некоторый тупиковый тест таблицы Т. В таблице Т выделяется подмножество столбцов с номерами $(i_1, i_2, \dots, i_\ell)$. Если для некоторого $S_j = \{t_{ji}\}$ ($j = 1, 2, \dots, m$) выполняются равенства $\{t_{ji_i} = t_{ji_k}\}$ для $i_k \in \{i_1, i_2, \dots, i_\ell\}$, то считают, что выбранный тест "голосует" за отнесение объектов S_j к j -му эталону.

В результате перебора всех тупиковых тестов каждый из m эталонов получает определенное количество голосов N_j , $j = 1, \dots, m$.

Естественно отнести объект S_j к классу К, если

$$N_K = \max \{N_j\}$$

Так как основные затраты машинного времени при реализации программы П2 попрежнему приходится на перебор тупиковых тестов, оказывается выгодным обрабатывать сразу группу объектов $\{S_j\}$. Данные процедуры реализуются программой П4. Процедура голосования по тесторам [7, 8] отличается от "голосования по тестам" лишь тем, что вместо тупиковых тестов используются тупиковые тесторы. Для этого предназначена программа П5.

Добавление. На базе программы П5 может быть также реализован алгоритм "голосование по тупиковым пакетам" [6]. Согласно данной работе пакет (i_1, \dots, i_ℓ) таблицы Т голосует за отнесение строки S_j (пробы) к таблице Т, если при добавлении строки S_j к Т, набор (i_1, \dots, i_ℓ) останется пакетом в полученной таким образом таблице. Аналогично проводится голосование и для "0-пакетов". Обозначим $F(S_j, T)$ - число голосов, поданных за отнесение пробы S_j к Т. Тогда "1-пакетная" и "0-пакетная" мера строки равняются соответственно:

$$B(s) = \frac{F(S_j, T)}{F(T)} \quad , \quad \alpha \quad \bar{B}(s) = \frac{F(S_j, \bar{T})}{F(\bar{T})}$$

Кроме того, с помощью П5 можно остраивать Q-тесты и производить голосование по ним.

§ 4. Вычисление тестовых и тесторных параметров с учетом длин тупиковых тестов и тесторов.

Длиной тупикового теста (тестора) называется число составляющих его столбцов. Учет распределения длин тестов и тесторов, как показывает практика, улучшает результаты распознавания. Важность учета распределения длин тупиковых тестов обосновывается в [9]. Основ-

ную роль в задачах тестового распознавания и классификации играют следующие величины, разработанные на тестовой основе [1] :

а) $P_i = \frac{K_i}{K}$ - информационный вес i -го столбца, где K - общее число тупиковых тестов, а K_i - нагрузка i -го столбца, или число вхождений i -го столбца в тупиковые тесты;

б) $J(s) = \sum_{x_i=1} P_i + \frac{1}{2} \sum_{x_i=-1} P_i$ - информационный вес строки $S = (x_1, x_2, \dots, x_n)$;

в) $P_i^* = \frac{K_i^*}{K^*}$ - тесторный информационный вес i -го столбца, где K^* - число всех тупиковых тесторов, а K_i^* - число их с участием i -го столбца.

Как видно из а) - в), информация о длинах тупиковых тестов или тесторов в операциях вычисления P_i , $J(s)$, P_i^* не участвует. Это является результатом того, что любые два тупиковые тестора (теста) считаются равноправными. Поскольку тест можно представить как частный случай тестора (для однострочных таблиц), в дальнейшем все формулировки проводятся для тесторов, но справедливы и для тестов.

Указанная выше основа а)б),в) сохраняется, но в дополнение к ней теперь предлагается некоторое расчленение веса столбцов, учитывающее длины, а именно: для j -го столбца определим (сравните [9]):

$$R_i^* = P_i^* \cdot \rho_i^* , \quad Q_i^* = P_i^* - R_i^* ,$$

где $\rho_i^* = 1 - \frac{\ell_i^*}{\ell_{max}^*}$, ℓ_i^* - средняя длине тупиковых тесторов с участием i -го столбца, а ℓ_{max}^* максимальное из значений ℓ_i .

В связи с этим расчленением и запросами уточнения веса столбцов возникает потребность в изучении распределения длин тупиковых тесторов. Задача эта решается вычислением количеств $K^{\ell*}$ тупиковых тесторов длины ℓ и чисел $K_i^{\ell*}$ тупиковых тесторов длины ℓ с участием i -го столбца. Остальные тесторные характеристики вычисляются по формулам:

$$K^* = \sum_{\ell \geq 1} K^{\ell*} , \quad K_i^* = \sum_{\ell \geq 1} K_i^{\ell*} ,$$

$$\bar{L}^* = \frac{1}{K^*} \sum_{\ell \geq 1} \ell \cdot K^{\ell*} \quad - \text{средняя длина тупикового тестора,}$$

$$\ell_i^* = \frac{1}{K_i^*} \sum_{\ell \geq 1} \ell \cdot K_i^{\ell*} \quad - \text{средняя длина тупиковых тесторов с } i\text{-м столбцом,}$$

$$\rho_i^* = 1 - \frac{\ell_i^*}{\ell_{max}^*} \quad - \text{коэффициент различия } i\text{-го столбца,}$$

$R_i^* = P_i^* \cdot \rho_i^*$ - тесторный различающий вес столбца,

$Q_i^* = P_i^* - R_i^*$ - тесторный отождествляющий вес i -го столбца

Характеристики следующие:

$J^*(S) = \sum_{x_i=1} P_i^* + \frac{1}{2} \sum_{x_i=-1} P_i^*$ - тесторный информационный вес строки S

$R^*(S) = \sum_{x_i=1} R_i^* + \frac{1}{2} \sum_{x_i=-1} R_i^*$ - тесторный различающий вес строки S ,

$Q^*(S) = J^*(S) - R^*(S)$ - тесторный отождествляющий вес строки S .

- употребляются обычно тогда, когда тесторы вырождаются в тесты, то есть являются тесторами для I-строчных таблиц.

Указанные выше величины вычисляются программой П6.

§ 5. Повторяемость столбцов и вычисление тестовых параметров для пятистрочных таблиц (П7)

Тестовые методы решения задачи сравнительного изучения уникальных геологических объектов, охарактеризованных значительным числом (до 400) логических признаков, требуют подсчета числа всех тупиковых P-тестов и Q-тестов, в соответствующих таблицах. По программе П7 "Состав" этот подсчет проводится для 5-строчных таблиц с произвольным числом столбцов. Алгоритм опирается на следующую теорему для таблиц с фиксированным числом строк [10, 11].

ТЕОРЕМА. $K = \sum_{t \in \tau} \Pi(t)$, $K^Q = \sum_{t \in \theta} \Pi(t)$ (I)

Здесь K - число всех тупиковых тестов в таблице T , K^Q - число всех тупиковых Q-тестов в таблице T , а τ и θ обозначают множества тупиковых тестов и Q-тестов в таблице T^* , в которой все столбцы бесповторны и таковы, что для каждого столбца таблицы T найдется либо равный, либо зеркальный столбец таблицы T^* ; причем $\Pi(t)$ определяется следующим образом: если $t = (i_1, i_2, \dots, i_q)$, то $\Pi(t) = z_{i_1} \cdot z_{i_2} \cdot \dots \cdot z_{i_q}$, где z_{i_q} суммарное число равных и зеркальных к i_q -му столбцов из T , где $q = 1, \dots, 5$.

Если в качестве T^* взять максимальную возможную таблицу из бесповторных столбцов, то по формулам (I) можно выполнить подсчет

ж) включая и столбец i_q .

K, K^a для любой таблицы T с тем же числом строк, что и T^* .
 В качестве T^* в программе "Состав" используется таблица

$$T_{max}^5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Для вычисления информационных весов признаков P_i, Q_i используются формулы:

$$P_i = \frac{K_i}{K}, \quad Q_i = \frac{K_i^a}{K^a}, \quad \text{где}$$

K_i - число тупиковых тестов с участием i -го столбца, а K_i^a - число всех тупиковых Q -тестов с наличием i -го столбца. Числа K_i, K_i^a находятся по формулам

$$K_i = K - K(i), \quad K_i^a = K^a - K^a(i),$$

где $K(i), K^a(i)$ - числа тупиковых тестов и соответственно Q -тестов в таблице $T(i)$, которая получается из T в результате удаления i -го столбца. Вычисление чисел $K(i), K^a(i)$ для таблиц $T(i)$ производится по формулам (1).

Л и т е р а т у р а

1. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. О математических принципах классификации предметов и явлений. Сб. "Дискретный анализ", изд. "Наука", Новосибирск, вып.7, 1966, с.3.
2. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. Об одном принципе классификации и прогноза геологических объектов и явлений. Геология и геофизика, № 5, 1968.
3. Смертин Е.А. Q -тесты в задачах тестового распознавания. Сб. "Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии", Новосибирск, 1973.
4. Слуцкая Т.Л. Алгоритмы вычисления информационных весов признаков. Сб. "Дискретный анализ", изд. "Наука", Новосибирск, вып.12, 1968, с. 6.

5. Чегис И.А., Яблонский С.В. Логические способы контроля работы электрических схем. Труды математического института им. В.А.Стеклова, 1958, 51, с.270.
6. Красавчиков В.О. Модификация тестового подхода к анализу таблиц описаний на основе понятия пакета. В кн."Дискретный анализ",Новосибирск, 1974, вып.26, с.36-60.
7. Мацак А.П. Проблема обучения по малым выборкам при геологическом прогнозировании. Известия АН КазССР, сер.физ-мат.,№ 5,1969.
8. Бугаец А.Н., Дворниченко Г.К., Мацак А.П., Серова Л.Л. Алгоритмы и программы решения геологических задач на ЭВМ "Минск-2" и БЭСМ-3М", Алма-Ата, Каз.научн.-исслед.ин-т минерального сырья", вып.2, 1969.
9. Дмитриев А.Н. Использование длин тупиковых тестов при обработке таблиц. Сб."Дискретный анализ", изд."Наука", Новосибирск, вып. 17, 1970, с.18.
10. Дмитриев А.Н., Смертин Е.А. Связь тестовых параметров таблиц с повторяемостью столбцов. Сб."Всесоюзная конференция по проблемам теоретической кибернетики",Новосибирск,тезисы,1969.
11. Дмитриев А.Н., Смертин Е.А. Алгоритмы вычисления тестовых параметров бинарных таблиц в задачах распознавания. Сб."Алгоритмы и программы решения геологических задач". Министерство Геологии СССР, Алма-Ата, вып.3, 1970.
12. Операционная система ДИСПАК для БЭСМ-6. Институт прикладной математики АН СССР, М.,1973.
13. Ляшенко В.Ф. Программирование для ЦВМ с системой команд типа М-220, М., "Сов.радио",1974.
14. Система пакетной обработки задач на ЭВМ М-220. АС- α -М-220. Отчет ВЦ СО АН СССР, Новосибирск,1972.

ПРОГРАММА ПІ

"Вычисление информационных весов признаков для ЭВМ типа М-220" составлена в машинных кодах.

Слуцкая Т.Л.

Назначение.

Программа ПІ предназначена для изучения важности признаков и объектов с помощью тестов [5]. Она вычисляет информационные веса признаков P_i и объектов $J(s)$ [1].

Инструкция к пользованию

I. Ограничения на исходную таблицу Т: а) число столбцов в таблице $n \leq 36$, число строк $m \leq 64$, однако, как правило, m предполагается значительно меньшим; б) каждый столбец исходной таблицы содержит хотя бы одну единицу; в) никакие две строки исходной таблицы не должны совпадать (следует иметь в виду, что прочерк считается совпадающим и с нулем, и с единицей). Останов в ячейке 0063 означает невыполнение этого условия.

II. Контрольными ячейками, позволяющими следить за счетом, являются:

ячейка 0010 - в ней находится набор \bar{X} , в процессе счета содержимое этой ячейки уменьшается от кода, содержащего единицы в первых n разрядах до нулевого кода.

Ячейка 0005 - в ней по второму адресу стоит число накопленных тупиковых тестов.

Останов в ячейке 0135 означает, что не хватает места для тупиковых тестов. При передаче управления в ячейку 0016 накопленные тесты обрабатываются.

Программный останов в ячейке 0107. Для останова по времени следует набрать останов по КРА в ячейке 0134 и передать управление в ячейку 0016.

III. Исходные данные представляют собой таблицу вида:

$$T = \{t_{ij}\}, \quad t_{ij} \in \{0, 1, -\}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m$$

N об	n_p	I_1	I_2	...	I_n
S_1		t_{11}	t_{21}	...	t_{n1}
S_2		t_{12}	t_{22}	...	t_{n2}
⋮		⋮	⋮	⋮	⋮
⋮		⋮	⋮	⋮	⋮
S_m		t_{1m}	t_{2m}	...	t_{nm}

Строки таблицы кодируются восьмеричными цифрами и записываются в адресной части ячеек, начиная с 7700 (в коде операции ставятся нули).

Если в строке есть прочерки, они заменяются нулями и строка записывается как обычно; непосредственно за ней вводится дополнительная строка с единицами в разрядах, соответствующих прочеркам, и с нулями — в остальных. Признаком дополнительной строки служит единица в 45-ом разряде, то есть КОП-400.

В ячейку 7677 заносится код:

0 00 $m'n$

где m' — длина таблицы (с учетом дополнительных строк), n — число столбцов.

Пример. Для таблицы

	1	2	3	4	5	6
1	0	1	1	0	0	-
2	0	1	0	1	1	0
3	1	0	-	-	1	1

информация имеет вид:

K	КОП	А I	А II	А III	
		7677	-	-	КА
	000	0005	0006	-	
	000	3000	-	-	
	400	0100	-	-	
	000	2600	-	-	
	000	4300	-	-	
	400	1400	-	-	

Перфокарты с информацией ставятся в начало колоды перед перфокартой № I после паспорта (шифр, время, инструкция, $K\Sigma$). Паспорт

перфорируется на КУ-3.

IV. Печать результатов.

После получения всех тупиковых тестов (или по передаче управления в ячейку 0016) печатаются следующие результаты:

1) (не основные) таблица T^* ; "след" таблицы T_2 , по которому столбцам и строчкам возвращаются их исходные порядковые номера; тестовые наборы в виде восьмеричных цифр в адресной части ячеек (тестовые наборы соответствуют порядку столбцов, принятому в T^*);

2) основные результаты печатаются 4-мя отдельными массивами (с тестами печати): (см. § I):

а) информационные веса признаков

$$P_i = \frac{K_i}{K}, \quad i = 1, 2, \dots, n,$$

где K - число всех тупиковых тестов;

K_i - число тупиковых тестов, в которые вошел i -ый признак; в начале массива P_i печатается число K ;

б) величины

$$P_i' = \frac{\lambda_i}{m} \cdot P_i$$

где λ_i - число единиц в i -ом столбце;

m - число строк исходной таблицы;

Обозначим через $J(S_j)$ и $J'(S_j)$ следующие величины:

$$J(S_j) = \sum_{i=1}^n |t_{ij}| \cdot P_i$$

$$J'(S_j) = \sum_{i=1}^n |t_{ij}| \cdot P_i'$$

Эти величины вычисляются и печатаются для всех строк (двумя отдельными массивами). Для приведенного образца результаты счета сведены в таблицу

№	1	2	3	4	5	6
P_i	0,43	0,43	0,28	0,28	0,43	0,14
P_i'	0,14	0,28	0,09	0,09	0,28	0,05

№	$J(S)$	$J'(S)$
1	0,786	0,405
2	1,143	0,666
3	1,286	0,571

Программа Q -тесты

"Вычисление Q -тестовых информационных весов признаков",
которая является производной ПІ, модифицирована Смертиным
Е.А.

В теле программы меняется перфокарта I5^a, которая принимает вид:

π	КОП	A1	A2	A3	
4	-	0024	-	-	KA
-	I5	0002	0020	0002	
4	72	-	00I2	00I3	
-	-	0050	-	-	KA
0	I6	005I	0024	0026	

Печать результатов идентична с ПІ при соответствующих переобозначениях:

K - число Q -тестов

$Q_i, Q'_i, Q(s), Q'(s)$ - Q -тестовые веса столбцов и строк.

ТЕКСТ ПРОГРАММЫ ПИ В МАШИННЫХ КОДАХ

	0 00 0001 0000 0000	КА	0110	0 01 0001 0000 0001	№6
	0 56 0000 7500 0000	№1		0 55 0001 0161 0001	
	0 00 0030 0000 0000	КА		I 51 0044 0110 0001	
0030	0 52 0001 0000 0012	№2		I 12 0000 0100 0000	
	I 32 0000 7517 0000			4 72 0000 0100 0002	
	4 00 7700 0000 0001			0 34 0162 0002 0002	
	4 55 7701 0077 0000			0 13 0163 0002 0113	
	6 52 0001 0001 0030		I20	4 52 7777 0000 0011	
	0 36 0076 0040 0010			0 56 0021 0100 0010	
	4 15 7700 0076 0010			0 72 0000 0011 0100	
	6 52 0001 0001 0030			4 75 0200 0015 0015	
40	4 15 7700 0001 0002			4 15 0200 0075 0202	
	0 55 0002 0010 0002			0 13 0123 0162 0123	№7
	4 55 7701 0077 0000			I 00 0160 0000 0200	
	I 11 0000 0050 0001			0 15 0123 0164 0000	
44	0 36 0075 0030 0015	№3	I30	0 76 0163 0100 0113	
	4 15 7700 0076 0011			0 12 0000 0000 0000	
	0 55 0002 0011 0002			I 00 0000 0000 0200	
	2 52 0000 0001 0000			I 12 0000 0131 0001	
50	4 72 0000 0012 0013			0 52 0000 0000 0000	
	4 55 0177 0002 0003			4 55 0202 0010 0001	
	4 15 0177 0003 0000			6 52 0001 0001 0133	
	0 36 0000 0064 0000			0 01 0001 0000 0002	
	0 15 0003 0002 0000			0 34 0002 0165 0002	
	I 71 0002 0051 7777		I40	0 36 0000 0152 0000	№8
	0 36 0000 0066 0000			5 75 0177 0166 0177	
	4 72 0000 0012 0000			4 55 0202 0002 0000	
0060	6 52 0001 0001 0012	№4		I 11 0000 0142 0001	
	I 15 0002 0000 0177			0 76 0000 0141 0000	
	0 76 0000 0064 0000			0 34 0167 0166 0166	
	0 17 0030 0013 0000			0 72 0000 0133 0000	
	0 72 0000 0013 0000		I50	0 15 0001 0002 0001	
	0 56 0000 0040 0000			0 73 0170 0136 0170	
	4 72 0000 0012 0014			0 56 0000 0136 0000	
	4 00 0177 0000 0004			4 15 0201 0076 0005	
70	4 72 7777 0014 0012			0 55 0005 0010 0010	
	I 00 0004 0000 0200		I54	3 13 0170 0201 0201	№9
	I 32 0001 0051 0000			0 55 0010 0076 0000	
	0 56 0000 0057 0000			I 51 0000 0134 0000	
74	0 00 0000 0000 0000	№5	I60	0 00 0000 0000 0000	
	I 00 0000 0000 0000			7 77 7777 7777 7777	
	0 00 7777 7777 7777			I 77 3777 7777 7777	
	4 00 0000 0000 0000			I 14 0000 0000 0001	
0100	0 52 0000 0000 0000			I 12 0000 0100 0000	
	4 15 0200 0160 0000			4 15 0200 0075 0202	
	I 11 0000 0101 0001			0 00 4000 0000 0000	
	4 75 0177 0015 0001			I 00 4000 0000 0000	
	0 62 0001 0021 0001			0 77 7777 7777 7777	
	6 52 0000 0000 0100			0 00 7500 0000 0000	№10
	0 32 0001 0121 0000				КА
	0 76 0000 0113 0000				

7500	0 55	7550	7677	0001	№I1						
	0 13	0031	0001	0031		0 00	0063	0000	0000	№I5a	
	0 13	0043	0001	0043		0 77	0030	0013	0000		
	0 13	7507	0001	7507							
	0 52	0000	0000	0000		0 00	0000	0000	0000	№Σ i	
	0 00	0000	0000	0020						№I6	
	2 75	0020	7700	0020	0031	0 00	0021	0000	0010	№A	
	I 12	7777	7506	0001		0 00	0000	0000	0007		
10	0 13	0075	0020	0021		0 52	0000	0000	0005		
	0 74	7551	7677	0001		4 55	0200	0010	0001		
	0 53	0075	0001	0170		4 15	0200	0001	0000		
	0 73	7552	0001	0002		I 51	7777	0036	0001		
7514	0 34	0002	0165	0022	№I2	0 76	0133	0055	0004		
	0 13	0075	0022	0023		4 73	0201	0100	0000		
	0 56	0076	0030	0200		4 76	0201	0051	0003		
	0 34	0162	0012	0012							
20	0 13	0102	0012	0102		0044	2 34	0102	0177	0001	№I7
	0 13	0106	0012	0106			0 55	0001	0100	0001	
	0 13	0132	0012	0132			4 72	0000	0001	0035	
	0 13	0134	0012	0134			2 75	0007	0137	0007	
	0 13	0143	0012	0143		50	0 56	0007	0035	0011	
	0 13	0142	0012	0142			0 34	0003	0023	0004	
	0 13	0152	0012	0152			0 62	0010	0004	0010	
	0 13	0156	0012	0156			0 71	0000	0036	0000	
7530	0 34	7553	0012	0001	№I3		0 56	0000	0016	0000	
	0 34	7553	0001	0002			0 15	0010	0020	0012	
	0 13	0123	0002	0123			0 15	0007	0000	0000	
	0 13	0123	0002	0164			0 36	0012	0136	0013	
	0 13	0154	0001	0154		0060	0 33	0131	0011	0011	№I8
	0 13	0154	0002	0154			0 01	0011	0000	0011	
	0 00	0170	0000	0025			0 73	0131	0011	0001	
	0 00	0012	0000	0026			0 74	0106	0001	0001	
40	0 16	7541	0100	0157			0 13	0012	0001	0013	
	0 10	0030	7542	0000			0 15	0011	0107	0000	
	0 13	0040	0026	0040			0 76	0101	0110	0123	
	0 13	0042	0026	0042			0 56	0000	0052	0007	
7544	0 13	0043	0026	0043	№I4	70	0 33	0131	0011	0011	
	0 13	0026	0026	0001			0 01	0011	0000	0011	
	0 13	0112	0001	0112			0 73	0131	0011	0001	
	0 56	0000	7554	0000			0 74	0106	0001	0001	
50	0 00	7777	0000	0000		0074	0 13	0012	0001	0012	№I9
	I 30	0000	0000	0000			0 02	0015	0012	0000	
	I 01	0000	0000	0000			0 76	0015	0060	0013	
	0 64	0000	0000	0000			0 56	0104	0110	0123	
	0 13	0113	0001	0113		0100	I 05	0000	0037	0000	
	0 14	0050	0001	0001			0 76	0013	0070	0012	
	0 13	0111	0001	0111			0 00	0000	0000	0000	
	0 14	0105	0023	0133			0 00	0000	0000	0000	
7560	0 13	0032	0001	0032	№I5		0 76	0000	0075	0000	
	0 14	0050	0026	0002			I 01	0000	0000	0000	
	0 13	0031	0002	0031			0 00	0000	0000	0000	
	0 14	0077	0022	0103			I 00	4000	0000	0000	
	0 73	0130	0025	0106		0110	0 72	0000	0005	0000	№20
	0 73	0025	0132	0102			I 56	0013	0134	0204	
	0 50	0500	0000	0300			4 55	0203	0013	0001	
	0 70	0001	0000	0000			4 15	0203	0001	0000	
7570	0 56	0000	0031	0000			I 71	0002	0112	7777	

	0 36 0I03 0II7 0002		0 00 0030 0000 0000	№26
	0 13 0005 0II3 0005			κA
	0 34 0I05 0002 0002		0030 0 13 0064 0026 0064	№27
120	0 55 00I3 0002 0000		0 13 0064 0026 0064	
	0 36 0000 0II7 0000		0 13 0050 0026 0050	
	0 13 00I3 0002 00I5		0 55 0005 0I22 000I	
	0 00 0000 0000 0000		0 13 0I23 000I 0005	
			0 0I 0005 0000 0I30	
0I24	0 56 0000 00I6 0000 №2I		0 00 0000 0000 0003	
	0 15 0020 00I5 00I0		0 52 0000 0000 00I0	
	0 32 0000 0036 0000	40	I 50 0500 0000 0I70	
	0 00 7777 7777 7760		4 70 0200 0000 0000	
130	2 00 0000 0000 0000		0 33 000I 0I20 000I	
	I 00 0000 0000 0000		I 7I 0000 0040 0770	
	0 30 0000 0000 0000			
	0 00 0000 0000 0000		0444 0 52 0000 0000 0000	№28
	I 12 7200 0II2 0000		I 00 0000 0000 0I3I	
	0 77 00I0 00II 0000		I 12 0044 0045 000I	
	0 56 0I37 0II0 0I23		0 13 0050 0026 0II7	
	0 76 0000 0I25 0000	50	0 55 0202 0020 000I	
			0 15 0050 0II7 0000	
0I40	I 00 5252 5252 5250 №22		0 36 0000 0075 0004	
	I 00 63I4 63I4 63I0		0 13 0050 0I2I 0050	
	I 00 42I0 42I0 42I0		0 15 0020 000I 0020	
	I 00 74I7 0360 74I0		0 13 0I24 000I 000I	
	I 00 50I2 0240 50I0		0 0I 000I 0000 000I	
	I 00 60I4 0300 60I0		0 55 0I27 000I 000I	
	I 00 40I0 0200 40I0			
	I 00 7760 0377 00I0	0060	0 36 0000 0050 0006	№29
0I50	I 00 5240 0252 00I0		0 73 0I24 000I 0002	
	I 00 6300 03I4 00I0		0 54 0050 0002 0002	
	I 00 4200 02I0 00I0		0 72 0000 0005 0000	
	I 00 7400 0360 00I0		4 55 0203 0I25 0000	
			0 36 0000 0067 0000	
0I54	I 00 5000 0240 00I0 №23		0 0I 0I26 0003 0003	
	I 00 6000 0300 00I0		I 32 0002 0064 7777	
	I 00 4000 0200 00I0	70	0 72 0000 0002 0000	
	I 00 7777 7400 00I0		I 04 0003 0I30 0I3I	
160	I 00 5252 5000 00I0		4 0I 0I3I 00I0 00I0	
	I 00 63I4 6000 00I0		0 14 0077 0I25 0I25	
	I 00 42I0 4000 00I0			
	I 00 74I7 0000 00I0	0074	0 56 0000 0056 0003	№30
	I 00 50I2 0000 00I0		0 12 0000 0000 0000	
	I 00 60I4 0000 00I0		3 75 0I24 7677 7077	
	I 00 40I0 0000 00I0		I 12 0I00 0076 000I	
	I 00 7760 0000 00I0	0I00	0 50 00I3 0000 7767	
			0 70 7500 0I00 0000	
0I70	I 00 5240 0000 00I0 №24		0 16 0I03 750I 76I0	
	I 00 6300 0000 00I0		0 52 0I30 0027 0200	
	I 00 4200 0000 00I0		0 00 0000 0000 0000	
	I 00 7400 0000 00I0		0 56 0000 0I15 0000	
	I 00 5000 0000 00I0		0 00 0000 0000 0000	
	I 00 6000 0000 00I0		0 00 0000 0000 0000	
	I 00 4000 0000 00I0			
		№25	0I00 0 00 0000 0000 0000	№3I
	0 00 00I6 0000 0000	κA	0 00 0000 0000 0000	
	0 10 0030 00I7 0000		0 00 0000 0000 0000	
	0 56 0005 0030 000I		0 00 0000 0000 0000	
			0 00 0000 0000 0000	
			0 10 0030 0II6 0000	
	3 37 I456 6470 0507	κ Σ 2		

0I24	I 00 0000 0000 0000	№32	0I10	4 00 0000 0000 0000	№38
	0 00 4000 0000 0000			0 00 4000 0000 0000	
	I 0I 4000 0000 0000			0 54 0050 0002 0002	
	I 77 3777 7777 7777			0 55 0002 0I15 0002	
				0 56 0000 0054 0000	
	2 I6 .7652 3I33 253I	κ Σ 3		0 00 0000 0077 0000	
		№33			№39
	0 00 0030 0000 0000	κA		0 00 0I07 0000 0000	κA
				3 77 0000 0000 0000	
0030	0 I4 0064 7077 000I	№34		6 II 4205 004I I533	κ Σ
	0 72 0000 000I 0002				
	4 55 7077 0I10 0000				
	0 76 0000 0040 0000				
	0 0I 0I26 0004 0004				
	4 55 7077 0I1I 0000				
	0 36 0000 0040 0000				
	0 0I 0I26 0002 0002				
40	I 32 0002 0032 7777				
	0 52 0000 0000 0000				
	0 04 0002 0004 0002				
	5 05 0I3I 0002 020I				
0044	6 52 000I 000I 004I	№35			
	4 0I 0200 0006 0006				
	0 I4 0077 0I1I 0I1I				
	0 76 0000 003I 0004				
50	0 32 0000 0075 7777				
	5 0I 7I00 0000 7I00				
	2 73 0I24 7I00 0002				
	0 56 0000 0I12 0000				
	3 55 0I27 7I00 7I00				
	0 36 0000 0062 0000				
	0 72 0000 0002 0000				
	4 0I 0I3I 0005 0005				
0060	4 0I 020I 0007 0007	№36			
	0 56 0000 0050 0000				
	0 72 0000 0050 0000				
	I 00 0005 0000 0250				
	I 00 0007 0000 0350				
	0 55 0003 0I10 0000				
	0 36 0000 0075 0000				
	2 06 0077 0250 0005				
70	2 06 0077 0350 0007				
	3 0I 0005 0247 0247				
	3 0I 0070 0347 0347				
	I 00 0000 0000 0250				
0074	I 00 0000 0000 0350	№37			
	6 52 000I 000I 0050				
	0 00 0000 0000 0005				
	4 I5 7I00 0I24 0003				
0I00	0 76 0000 005I 0007				
	0 I6 0I02 750I 76I0				
	0 52 0200 0027 0252				
	0 I6 0I04 750I 76I0				
	I 52 0250 0027 0252				
	0 I6 0I06 750I 76I0				
	I 52 0350 0027 0352				
	0 I7 0000 0000 0000				

ПРОГРАММА П2

"Вычисление информационных весов признаков" для ЭВМ
БЭСМ-6, составлена в машинных кодах.

Слуцкая Т.Л.

Назначение.

Программа П2 предназначена для вычисления информационных весов признаков (P_i), участвующих в табличном описании множества, представленного в виде $T(m \times n)$. П2 представляет собой модификацию P_i , приспособленную для обработки информации на БЭСМ-6.

Описание программы.

Собственно программа занимает ячейки памяти I2070-I2377. Ниже приводится функциональное назначение отдельных участков (блоков) программы.

Блок 1 (I2070-I2076) - начальные засылки

$0 \dots 0 \longrightarrow$ в массив информ.весов,

$1 \dots 1 \longrightarrow$ в массив прочерков.

Блок 2 (I2076-I2104) - расшифровка информации о прочерках, если они есть; обращение к ОП-2 для расшифровки шкал.

Блок 3. (I2105-I2111) - попарное сравнение 2-х строк: j -ой с j' -ой ($j < j'$) с учетом информации о прочерках.

Блок 4 (I2112-I2126) - составление таблицы T - результатов сравнений с проверкой на поглощение; таблица T на каждом шаге состоит из попарно несравнимых строк; таблица укорачивается за счет переноса последних строк на место "лишних", если такие появятся по ходу составления T .

Блок 5 (I2127-I2143) - строки таблицы T переупорядочиваются в зависимости от количества единиц; первой ставится строка с минимальным количеством единиц, затем - с наименьшим числом единиц в оставшихся разрядах и т.д.

Блок 6 (I2144-I2161, I2264-I2270) - переупорядочивание столбцов в новой таблице так, чтобы единицы концентрировались в левом нижнем углу таблицы. Это достигается вынесением на первые места столбцов, имеющих единицы в первой строке и т.д. В ячейках I2000-I2060 - запоминается истинный порядок столбцов.

Блок 7 (I2161-I2165) - нахождение логической суммы всей таблицы для определения номера последней единицы.

Обращение к СП-I.

Блок 8 (I2I66-I2I72)- проверка очередного набора X (яч. I2073) на "тест-нет".

Блок 9 (I2I73-I2I76)- если X - не тест - образование нового X (обращение к СП-I);

Блок IO(I2I77) - проверка нового X на конец.

Блок II(I2200-I2205)- если X- тест, то проверка на тупиковость; образование очередного поднабора (опусканием одной из единиц) для проверки его на "тест-нет".

Блок I2(I2206-I2207)- если тест X не тупиковый- подготовка для образования нового X.

Блок I3(I22II-I22I7)- если X - тупиковый тест- обработка его: прибавление единицы к информационным числам столбцов-признаков, которые входят в X.

Блок I4(I2220-I2225)- СП-I - определение номера последней единицы в коде, находящемся в ячейке I2070, возврат по R IO, результат в виде адреса I2300+n в R7, где n - номер последней единицы.

Блок I5(I2226-I2232)- СП-2 - определение номера очередной единицы в коде по адресу I2076, возврат после каждой единицы по R IO, возврат после просмотра всего кода по адресу I220I+R I2 результат в виде, аналогичном СП-I.

Блок I6(I2240-I2245)- печать результатов.

Блок I7(I2252-I2260)- организация промежуточных печатей через каждые 2^{IO} тупиковых тестов.

Блок I8(I2374-I2377)- организация возможности прерывания и продолжения счета.

Блок I9(I230I-I2367)- константы.

Распределение памяти

I2500-I4477 - массив основной информации;

OIOO0-OIO27 - шкалы, указывающие строки с прочерками;

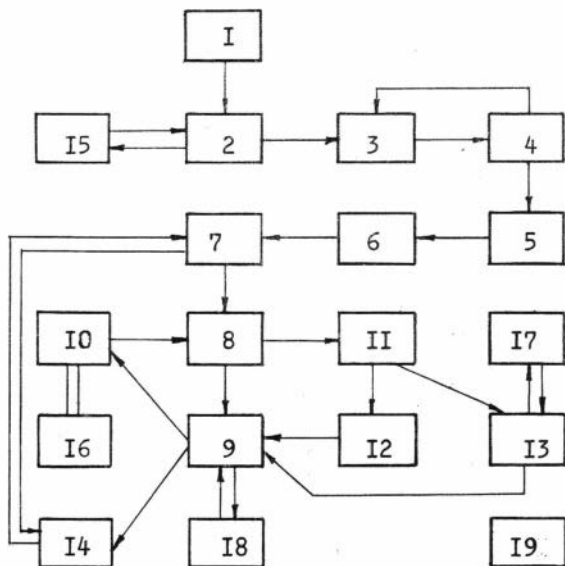
OIO30-O3027 - строки с единицами вместо прочерков;

Эти три массива имеют фиксированное начало и переменный конец, т.к. длина их зависит от количества объектов, поэтому конец массивов отмечается нулевым кодом. Эти массивы заполняются пользователем.

I4500-I6477- массив расшифрованной информации о прочерках.

I2070-I2377- программа

I2000-I2060-истинный порядок столбцов



ПРОГРАММА П2

Структурная блок-схема

I2400-I246I - информационные веса.

70000-77777 - таблицы результатов попарных сравнений строк;

в случае больших таблиц этот массив может быть легко увеличен за счет включения дополнительных листов памяти в паспорт, т.к. таблица заполняется снизу вверх и начало ее не фиксируется.

Инструкция к пользованию

I. Задание информации

Таблица, описывающая наличие (код "I"), отсутствие (код "O") или неопределенность ("прочерк") каких-либо признаков для группы объектов, кодируется восьмеричными цифрами и помещается в память машины по одному объекту в ячейку, начиная с адреса I2500.

После последнего объекта помещается нулевой код, свидетельствующий о конце таблицы (предполагается, что чисто нулевых строк таблица не содержит).

Порядок постановки перфокарт:

1. Паспорт.
2. Время.
3. E
4. Программа (0-35).
5. Информация.
6. E - конец.

Пример. Пусть таблица имеет вид:

IOI	OOO	OIO	III
OII	OII	IOO	OOI
IOO	ICO	OII	OIO

Эта таблица на бланке будет закодирована так:

B	00	OOI	2500
	00	OOO	0000
C	50	27	0000
	00	00	0000
C	33	4I	0000
	00	00	0000
C	44	32	0000
	00	00	0000

С	00	00	0000
	00	00	0000

Если таблица содержит прочерки, то указанные перфокарты составляются аналогично, причем прочерки полагаются в них нулями, но кроме этих перфокарт задается еще дополнительная информация о прочерках.

Начиная с ячейки 01000 помещаются шкалы, которые указывают номера строк с прочерками единицами в соответствующих разрядах, Список шкал также заканчивается нулевым кодом. Информация о самих прочерках помещается начиная с ячейки 01030, в виде восьмеричных кодов, единицы которых соответствуют прочеркам.

Пример. Таблица такова:

10-	0-1	001
010	110	110
-01	-00	101

Полная информация о ней имеет вид:

В	00	001	2500
	00	000	0000
С	41	10	0000
	00	00	0000
С	26	60	0000
	00	00	0000
С	10	50	0000
	00	00	0000
С	00	00	0000
	00	00	0000
В	00	000	1000
	00	000	0000
С	50	00	0000
	00	00	0000
С	00	00	0000
	00	00	0000
В	00	000	1030
	00	000	0000
С	12	00	0000
	00	00	0000
С	44	00	0000
	00	00	0000

С 00 00 0000
 00 00 0000

2. Ограничения.

Признаков объектов, т.е. столбцов, в таблице предполагается не больше 48; для строк отведено 1024 ячейки, однако, во всех просчитанных таблицах их было существенно меньше (не больше 50).

Предполагается, что таблица не содержит "сплошных" столбцов и нулевых строк (при сравнении "прочерк" считается совпадающим и с нулем и с единицей). Из одинаковых строк, если они имеются, автоматически учитывается лишь одна, остальные опускаются.

3. Так как достаточно большие задачи считаются иногда по несколько часов, введена промежуточная печать для контроля счета и ориентировочной оценки оставшейся части работы. Эта печать производится через каждые 1024 тупиковых теста и содержит

а) печать содержимого рабочих ячеек I2070-I2077;

б) K - общее число просмотренных тупиковых тестов;

в) K_i - число вхождений i -го признака в тупиковые тесты.

В ячейке I2075 содержится X - набор, пробегающий в процессе работы значения от 00...0 до 11...1 в n старших разрядах (n - число признаков). По значению этого набора можно приблизительно оценить оставшуюся часть работы.

5. В конце работы программы на печать выдаются значения:

K - число всех тупиковых тестов

$$P(i) = \frac{K_i}{K} \text{ - информационный вес } i\text{-го признака для } i = 1, 2, \dots, n.$$

Для приведенного выше контрольного примера результаты счета сведены в таблицу:

№	1	2	3	4	5	6	7	8	9
1	1	0	-	0	-	1	0	0	1
2	0	1	0	1	1	0	1	1	0
3	-	0	1	-	0	0	1	0	1
P_i	0,000	0,200	0,200	0,000	0,20	0,50	0,50	0,20	0,20

ТЕКСТ ПРОГРАММЫ П2 В МАШИНЫХ КОДАХ

Паспорт:

№№№№ — хххххх —

Лист — 0-1, 5-7, 35-37 —

Вход — I2070 —

Авост. —

Время — х хх хх —
 час мин сек

	00	00I	2070	0					
BC	00	000	0000						
I2070	0I	247	77I7	I		II	000	0076	3
	II	24I	2000		K	00	27I	2226	
	00	0I0	0000			0I	24I	2500	
K	II	000	0074		K	02	24I	2500	
	00	22I	246I			03	240	0000	
K	0I	000	0000		K	07	24I	2000	
	0I	37I	2072			0I	0I0	0000	
K	I2	240	0I67		K	02	0I2	000I	
	0I	247	600I			0I	0II	2000	
K	II	0I0	036I		I2IIO	K	02	0II	200I
	00	22I	6477			II	000	0070	
K	0I	000	0000		K	00	26I	2I23	
	0I	37I	2075	2		03	34I	2I2I	4
K	02	240	0000		K	03	0I0	0000	
	04	240	0000			II	0II	0070	
K	I0	3II	2I03		K	II	000	007I	
I2I00	02	0I0	I030			03	0I2	0000	
K	II	0I2	036I		K	00	26I	2I23	
	07	000	2I77			II	0I0	007I	
K	02	250	000I		K	II	0I2	0070	
	00	30I	2226			00	27I	2I20	
K	00	000	0000		K	04	0I0	0000	
	0I	250	000I			03	000	0000	
K	0I	0I0	0777		K	04	250	000I	

I2I20	K	03 37I 2II2	5			II 000 0070	8
	K	00 000 0000			K	II 010 036I	
		04 044 0005			K	0I 100 7777	
	K	04 257 7777			K	05 37I 2130	
		II 010 0070			K	04 044 0003	
	K	04 000 0000			K	00 010 0000	
		04 044 0003			K	03 000 0000	
	K	02 250 000I			K	03 37I 2I45	
		02 010 000I			K	06 250 000I	
	K	00 27I 2I07			K	II 010 036I	
		0I 250 000I			K	II 000 0070	
	K	0I 044 0002			K	II 010 0362	
		0I 010 000I	6			00 040 0005	9
	K	00 27I 2I07		I2I50	K	06 044 0002	
		II 010 036I			K	04 227 7777	
	K	II 000 0070			K	02 010 0000	
		II 010 0365			K	II 0II 0070	
I2I30	K	II 000 007I			K	II 000 007I	
		04 044 0003			K	II 012 0070	
	K	03 250 000I			K	00 30I 2264	
		03 110 7777			K	04 227 7777	
	K	II 0II 0070			K	02 010 0000	
		II 022 0364			K	II 020 007I	
	K	II 000 0072			K	05 036 0000	
		II 005 007I	7		K	02 115 7777	10
	K	00 26I 2I37			K	02 100 7777	
		II 010 0072			K	02 37I 2I54	
	K	II 000 007I			K	II 010 007I	
		03 044 000I			K	II 022 0362	
	K	04 044 0006		I2I60	K	II 000 0362	
		03 37I 2I32			K	06 37I 2I50	
	K	0I 110 7777			K	04 250 000I	
		04 220 0000			K	04 044 000I	
I2I40	K	05 100 7777			K	00 010 0000	
		II 0II 0070			K	0I 115 7777	
	K	II 012 0070			K	0I 37I 2I63	

		II	000	0070	II		I2	240	0000	I4
	K	I0	3II	2220			K	II	010	0075
		07	010	0000				II	000	0073
	K	00	30I	2374			K	00	30I	2174
		II	000	0073				00	30I	2206
	K	I2	240	0000		I2210	K	00	000	0000
		04	044	000I				00	010	0000
	K	00	000	0000			K	II	000	0074
		01	II0	7777				I2	240	0005
I2I70	K	II	0II	0073			K	II	010	0075
		I2	26I	2I73				II	000	0076
	K	0I	37I	2I70			K	00	30I	2252
		I2	30I	2200	I2			00	000	0000
	K	00	000	0000			K	I0	3II	2226
		0I	II0	7777				II	010	0363
	K	00	000	0000			K	07	237	7500
		II	000	0070				II	004	0400
	K	I0	3II	2220			K	07	237	7500
		II	010	0073				II	000	0400
	K	00	30I	2376			K	00	30I	22I4
		II	000	0073				00	000	0000
	K	II	0I2	0360		I2220	K	00	010	0000
		00	27I	2I67				II	000	0074
	K	00	30I	2240			K	II	010	0070
		II	010	0073	I3			II	023	0074
I2200	K	II	000	0075			K	II	000	0074
		II	000	0076				00	03I	0000
	K	00	010	0000			K	00	27I	2222
		II	000	0074				II	230	0074
	K	00	000	0000			K	07	24I	2300
		I2	240	0010				I0	300	0000
	K	I0	3II	2226			K	00	000	0000
		II	010	0075				II	010	0076
	K	07	012	0000			K	I2	26I	220I
		II	000	0073				II	023	0074
	K	00	30I	2I67			K	II	000	0074

		00	03I	0000	I7		II	010	0400	20
I2230	K	II	000	0076		K	II	005	0260	
		II	230	0074			00	27I	2233	
	K	07	24I	2300		K	II	010	0260	
		IO	300	0000			II	004	026I	
	K	00	000	0000		K	II	000	0260	
		07	240	0300			II	064	0250	
	K	00	30I	2215		K	00	000	0000	
		00	000	0000			II	064	0236	
	K	00	000	0000		K	00	000	0000	
		00	000	0000			00	30I	2233	
	K	00	000	0000		K	00	000	0000	
		00	00I	2400	I8		45	50	0000	2I
	K	00	00I	246I		I2260	C	00	00	0000
		03	00I	00I6			45	50	0000	
	K	II	022	0006			C	00	00	0000
I2240	K	0I	247	7720			00	00	0000	
		00	000	0000			C	00	00	0000
	K	00	22I	246I			00	00	0000	
		0I	010	0000			C	00	00	0000
	K	II	016	0400			II	000	0070	
		00	22I	246I			K	00	010	0000
	K	0I	000	0000			II	000	0067	
		0I	37I	224I			K	II	010	007I
		II	064	0236	I9		00	26I	2154	22
	K	00	000	0000			K	II	023	0067
		00	074	0000			II	000	0067	
	K	00	000	0000			K	07	000	000I
		II	064	0250			00	03I	0000	
	K	00	000	0000		I2270	K	07	37I	2266
		00	30I	2244			00	000	0000	
	K	00	000	0000			K	00	000	0000
		00	00I	2070			00	000	0000	
I2250	K	00	00I	2100			K	00	000	0000
		02	00I	0020			00	000	0000	
	K	IO	023	0004			K	00	000	0000

		00	00I	230I	23						
B		00	00I	2I0I			C	00	00	0I00	26
		40	00	0000				00	00	0000	
C		00	00	0000			C	00	00	0040	
		20	00	0000				00	00	0000	
C		00	00	0000			C	00	00	0020	
		I0	00	0000				00	00	0000	
C		00	00	0000			C	00	00	00I0	
		04	00	0000				00	00	0000	
C		00	00	0000			C	00	00	0004	
		02	00	0000				00	00	0000	
C		00	00	0000			C	00	00	0002	
		0I	00	0000	24			002	00	0000	
C		00	00	0000		I2330	C	00	00	000I	27
		00	40	0000				00	00	0000	
C		00	00	0000			C	00	00	0000	
		00	20	0000				40	00	0000	
C		00	00	0000			C	00	00	0000	
		00	I0	0000				20	00	0000	
C		00	00	0000			C	00	00	0000	
		00	04	0000				I0	00	0000	
C		00	00	0000			C	00	00	0000	
		00	02	0000				04	00	0000	
C		00	00	0000			C	00	00	0000	
		00	0I	0000	25			02	00	0000	
I23I4	C	00	00	0000			C	00	00	0000	28
		00	00	4000				0I	00	0000	
C		00	00	0000			C	00	00	0000	
		00	00	2000				00	40	0000	
C		00	00	0000		I2340	C	00	00	0000	
		00	00	I000				00	20	0000	
C		00	00	0000			C	00	00	0000	
		00	00	0400				00	I0	0000	
320	C	00	00	0000			C	00	00	0000	
		00	00	0200				00	04	0000	
C		00	00	0000			C	00	00	0000	
								00	02	0000	

I2344	C	00	00	0000	29		00	00	0000	32	
		00	01	0000		C	00	00	0060		
	C	00	00	0000			00	00	0000		
		00	00	4000		C	00	00	0000		
	C	00	00	0000			II	010	0367		
		00	00	2000	I2370	K	II	013	0366		
	C	00	00	0000			II	000	0367		
		00	00	1000		K	II	000	0074		
50	C	00	00	0000			00	307	2103		
		00	00	0400		K	00	000	0000		
	C	00	00	0000			00	000	0000		
		00	00	0200		K	00	000	0000		
	C	00	00	0000	30		II	000	0071	33	
		00	00	0100	I2374	K	II	010	0461		
	C	00	00	0000			00	301	2166		
		00	00	0040		K	00	000	0000		
	C	00	00	0000			07	013	0000		
		00	00	0020		K	II	000	0461		
	C	00	00	0000			00	301	2176		
		00	00	0010		K	00	000	0000		
	C	00	00	0000			00	000	0020	34	
		00	00	0004		B	00	000	0000		
	C	00	00	0000			00	301	2246		
		00	00	0002		K	00	000	0000		
		00	00	0000	31		B	00	000	1000	35
I2360	C	00	00	0001			00	000	0000		
		77	77	7777		C	00	00	0000		
	C	77	77	7777			00	00	0000		
		00	00	0000			00	00	0000		
	C	00	00	0100							
		40	50	0000							
	C	00	00	0000							
		64	00	0000							
	C	00	00	0000							
		64	00	0000							
	C	00	00	0100							

ПРОГРАММА ПЗ

"Вычисление тесторной информационной меры признаков" для ЭВМ БЭСМ-6, составлена в машинных кодах.

Слущкая Т.Л.

Назначение.

Программа ПЗ предназначена для вычисления тесторной меры признаков [I].

Описание программы.

Если заданы две или несколько таблиц T_1, T_2, \dots, T_K , то тупиковым тестором для них называется набор (i_1, i_2, \dots, i_r) столбцов такой, что никакие две подтаблицы среди подтаблиц T'_1, T'_2, \dots, T'_K , составленных из этих столбцов, не имеют одинаковых строк и никакой поднабор $(i_{21}, i_{22}, \dots, i_{2r})$, входящий в (i_1, i_2, \dots, i_r) , не обладает этим свойством [I].

Тесторный вес признака $P(i^*)$ принимается равным отношению K_i^*/K^* , где K^* - общее число тупиковых тесторов таблиц, K_i^* - число тупиковых тесторов, в которые входит i -й столбец.

Алгоритмы для нахождения всех тупиковых тесторов аналогичны алгоритмам для тупиковых тестов. Требование, чтобы подтаблицы отличались только друг от друга, сказывается на первой части работы программы, когда производится попарное сравнение строк. В пределах таблицы строки могут повторяться. Очевидно, для нахождения тесторов не следует сравнивать строки каждой таблицы между собой, а лишь со строками остальных таблиц. Это достигается добавлением к основной программе П2 дополнительного блока (ячейки I6500-I65I3), который организует попарное сравнение всех строк из разных таблиц и только их.

Инструкция к пользованию

1. Заменить в основной колоде П2 перфокарты с № 3 и с № 5 на соответствующие № 3т и № 5т, приведенные ниже (в тексте пр-мы)
2. Добавить в конец колоды п/к №№ 36,37,38, также приведенные ниже (в тексте программы).
3. Информация к программе задается в следующем виде:
 - а) все таблицы T_1, \dots, T_K кодируются подряд так, что к

последней строке таблицы T_i примыкает первая строка таблицы T_{i+1} .
 Признаком конца является нулевой код, которым заканчивается последняя таблица T_K ;

б) если в таблицах есть строки с прочерками, то информация о них задается в виде, указанном в инструкции к П2, причем шкалы составляются так, как будто бы строки всех таблиц пронумерованы подряд (см. пример);

в) задается информация о началах таблиц T_1, T_2, \dots, T_K в K ячейках, начиная с I2462, так что в разрядах I-15 ячейки I2462+ (i-1) стоит код $(I2500 + \sum_{z=1}^{i-1} m_z)$, где m_z - длина T_z ; в ячейку I2642+K помещается нулевой код;

г) общая длина таблиц не должна превышать I020 слов;

д) в случае, если в разных таблицах окажутся одинаковые строки, произойдет останов и в двух первых напечатанных словах (в разрядах I-15) будут стоять адреса совпадающих строк.

Порядок постановки перфокарт:

1. Паспорт.
2. Время.
3. E
4. Программа (0-38).
5. Информация.
6. E - конец.

Пример задания информации к П3

Пусть таблицы таковы:

T_1	00I	IOI	I-0	T_3	III	000	0IO
	-IO	00I	00I		0-0	I-I	IOI
	I00	II0	00I				
T_2	-IO	0IO	-II				
	I00	IOI	000				

Закодированная информация имеет вид:

B	00	00I	2500	I	00	00	0000	
	00	000	0000		45	00	0000	
C	I5	40	0000	C	00	00	0000	
	00	00	0000		70	20	0000	2
C	2I	30	0000	C	00	00	0000	
	00	00	0000		05	50	0000	
C	46	I0	0000	C	00	00	0000	
	00	00	0000		00	00	0000	
C	22	30	0000	C	00	00	0000	
					00	00	0000	

B	00	000	I000		C	22	00	0000	
	00	000	0000			00	00	0000	
C	64	40	0000		C	00	00	0000	
	00	00	0000			00	00	0000	
C	00	00	0000		B	00	00I	2462	4
	00	00	0000			00	000	0000	
B	00	000	I030	3	C	00	00	0000	
	00	000	0000			00	0I	2500	
	00	20	0000			00	00	0000	
C	00	00	0000		C	00	0I	2503	
	40	00	0000			00	00	0000	
C	00	00	0000		C	00	0I	2505	
	40	40	0000			00	00	0000	
C	00	00	0000		C	00	00	0000	

Для приведенного контрольного примера результаты сведены в таблицу.

T1

№	I	2	3	4	5	6	7	8	9
1	0	0	I	I	0	I	I	-	0
2	-	I	0	0	0	I	0	I	I
3	I	0	0	I	I	0	0	0	I

T2

№	I	2	3	4	5	6	7	8	9
1	-	I	0	0	I	0	-	I	I
2	I	0	0	I	0	I	0	0	0

T3

№	I	2	3	4	5	6	7	8	9
1	I	I	I	0	0	0	0	I	0
2	0	-	0	I	-	I	I	0	I

P_i^*									
	0,400								
	0,100								
	0,550								
	0,450								
	0,400								
	0,600								
	0,500								
	0,450								
	0,550								

Для вычисления отождествляющих весов признаков Q_i таблицы T с помощью ПЗ необходимо:

В качестве таблицы T_1 взять T , а в качестве T_2 — зеркальную к ней T зерк. Тогда Q -тесты соответствуют тесторам этих двух таблиц, а Q_i равняется $P_i^*[6]$.

Для вычисления пакетных величин A_i и \bar{A}_i таблицы T (I- и O-пакетные веса признаков) необходимо:

— в качестве таблицы T_1 взять T , а в качестве T_2 — таблицу $T'2$ — нулевую строку ($T'2$ — единичную строку). Тогда "I-пакеты" соответствуют тесторам таблиц $T_1, T'2$, а "O-пакеты" соответствуют тесторам таблиц $T_1, T'2$. Поэтому $A_i = P_i^{*'}$, а $\bar{A}_i = P_i^{*''}$, где $P_i^{*'}$ — тесторная мера i -го признака для $T_1, T'2$; $P_i^{*''}$ — тесторная мера i -го признака.

При подготовке информации следует добавить к обеим таблицам в I-м случае — единичный, а во втором — нулевой отождествляющий столбец, для выполнения одного из условий данной программы, предусматривающего отсутствие строки, состоящей из сплошных нулей и единиц.

ТЕКСТ ПРОГРАММЫ ПЗ В МАШИНЫХ КОДАХ

Паспорт :

Шифр ⊂ xxxxx -

Лист ⊂ 0-I,5-7,35-37 -

Вход ⊂ I2070 -

Авост -

Время ⊂ х хх хх -

	BC	час	мин	сек						
		00	00I	2070	0					
		00	000	0000						
I2070	К	0I	247	77I7	I		II	000	0076	3т
		II	24I	2000		К	00	27I	2226	
		00	0I0	0000			03	240	0000	
	К	II	000	0074		К	07	24I	2000	
		00	22I	246I			00	30I	6500	
	К	0I	000	0000		К	00	000	0000	
		0I	37I	2072			0I	0I0	0000	
	К	I2	240	0I67		К	02	0I2	0000	
		0I	247	600I			0I	0II	2000	
	К	II	0I0	036I		I2IIO	К	02	0II	2000
		00	22I	6477			II	000	0070	
	К	0I	000	0000		К	00	26I	65II	
		0I	37I	2075	2		03	34I	2I2I	4
	К	02	240	0000		К	03	0I0	0000	
		04	240	0000			II	0II	0070	
	К	10	3II	2I03		К	II	000	007I	
		02	0I0	I030			03	0I2	0000	
I2I00	К	II	0I2	036I		К	00	26I	2I23	
		07	000	2I77			II	0I0	007I	
	К	02	250	000I		К	II	0I2	0070	
		00	30I	2226			00	27I	2I20	
	К	00	000	0000		К	04	0I0	0000	
		0I	250	000I			03	000	0000	
	К	0I	0I0	0777		К	04	250	000I	

I2I20	K	03	37I	2II2	5π		II	000	0070	8
		00	000	0000		K	II	0I0	036I	
		04	044	0005			0I	I00	7777	
	K	04	257	7777		K	05	37I	2I30	
		II	0I0	0070			04	044	0003	
	K	04	000	0000		K	00	0I0	0000	
		04	044	0003			03	000	0000	
	K	02	250	000I		K	03	37I	2I45	
		02	0I0	0000			06	250	000I	
	K	00	27I	2I07		K	II	0I0	036I	
		00	30I	6505			II	000	0070	
	K	00	000	0000		K	II	0I0	0362	
		0I	0I0	000I	6		00	040	0005	9
	K	00	27I	2I07		I2I50	K	06	044	0002
		II	0I0	036I			04	227	7777	
	K	II	000	0070		K	02	0I0	0000	
		II	0I0	0365			II	0II	0070	
I2I30	K	II	000	007I		K	II	000	007I	
		04	044	0003			II	0I2	0070	
	K	03	250	000I		K	00	30I	2264	
		03	IIO	7777			04	227	7777	
	K	II	0II	0070		K	02	0I0	0000	
		II	022	0364			II	020	007I	
	K	II	000	0072		K	05	036	0000	
		II	005	007I	7		02	II5	7777	IO
	K	00	26I	2I37		K	02	I00	7777	
		II	0I0	0072			02	37I	2I54	
	K	II	000	007I		K	II	0I0	007I	
		03	044	000I			II	022	0362	
	K	04	044	0006	I2I60	K	II	000	0362	
		03	37I	2I32			06	37I	2I50	
	K	0I	IIO	7777		K	04	250	000I	
		04	220	0000			04	044	000I	
I2I40	K	05	I00	7777		K	00	0I0	0000	
		II	0II	0070			0I	II5	7777	
	K	II	0I2	0070		K	0I	37I	2I63	

	K	II	000	0070	II		I2	240	0000	I4
		I0	3II	2220			K	II	0I0	0075
		07	0I0	0000				II	000	0073
	K	00	30I	2374			K	00	30I	2I74
		II	000	0073				00	30I	2206
	K	I2	240	0000	I22I0		K	00	000	0000
		04	044	000I				00	0I0	0000
	K	00	000	0000			K	II	000	0074
		0I	II0	7777				I2	240	0005
I2I70	K	II	0II	0073			K	II	0I0	0075
		I2	26I	2I73				II	000	0076
	K	0I	37I	2I70			K	00	30I	2252
		I2	30I	2200	I2			00	000	0000
	K	00	000	0000			K	I0	3II	2226
		0I	II0	7777				II	0I0	0363
	K	00	000	0000			K	07	237	7500
		II	000	0070				II	004	0400
	K	I0	3II	2220			K	07	237	7500
		II	0I0	0073				II	000	0400
	K	00	30I	2376			K	00	30I	22I4
		II	000	0073				00	000	0000
	K	II	0I2	0360	I2220		K	00	0I0	0000
		00	27I	2I67				II	000	0074
	K	00	30I	2240			K	II	0I0	0070
		II	0I0	0073	I3			II	023	0074
I2200	K	II	000	0075			K	II	000	0074
		II	000	0076				00	03I	0000
	K	00	0I0	0000			K	00	27I	2222
		II	000	0074				II	230	0074
	K	00	000	0000			K	07	24I	2300
		I22	240	00I0				I0	300	0000
	K	I0	3II	2226			K	00	000	0000
		II	0I0	0075				I0	0I0	0076
	K	07	0I2	0000			K	I2	26I	220I
		II	000	0073				II	023	0074
	K	00	30I	2I67			K	II	000	0074

		00	03I	0000	I7		II	010	0400	20
I2230	K	II	000	0076		K	II	005	0260	
		II	230	0074			00	27I	2233	
	K	07	24I	2300		K	II	010	0260	
		IO	300	0000			II	004	026I	
	K	00	000	0000		K	II	000	0260	
		07	240	0300			II	064	0250	
	K	00	30I	22I5		K	00	000	0000	
		00	000	0000			II	064	0236	
	K	00	000	0000		K	00	000	0000	
		00	000	0000			00	30I	2233	
	K	00	000	0000		K	00	000	0000	
		00	00I	2400	I8		45	50	0000	2I
	K	00	00I	246I		I2260	C	00	00	0000
		03	00I	00I6			45	50	0000	
	K	II	022	0006			C	00	00	0000
		0I	247	7720			00	00	0000	
I2240	K	00	000	0000			C	00	00	0000
		00	22I	246I			00	00	0000	
	K	0I	0I0	0000			C	00	00	0000
		II	0I6	0400			II	000	0070	
	K	00	22I	246I		K	00	0I0	0000	
		0I	000	0000			II	000	0067	
	K	0I	37I	224I		K	II	0I0	007I	
		II	064	0236	I9		00	26I	2I54	22
	K	00	000	0000		K	II	023	0067	
		00	074	0000			II	000	0067	
	K	00	000	0000		K	07	000	000I	
		II	064	0250			00	03I	0000	
	K	00	000	0000		I2270	K	07	37I	2266
		00	30I	2244			00	000	0000	
	K	00	000	0000		K	00	000	0000	
		00	00I	2070			00	000	0000	
I2250	K	00	00I	2I00		K	00	000	0000	
		02	00I	0020			00	000	0000	
	K	IO	023	0004		K	00	000	0000	

		00	00I	230I	23		00	00	0100	26	
	B	00	00I	210I			C	00	00	0000	
		40	00	0000				00	00	0040	
	C	00	00	0000			C	00	00	0000	
		20	00	0000				00	00	0020	
	C	00	00	0000			C	00	00	0000	
		I0	00	0000				00	00	00I0	
	C	00	00	0000			C	00	00	0000	
		04	00	0000				00	00	0004	
	C	00	00	0000			C	00	00	0000	
		02	00	0000				00	00	0002	
	C	00	00	0000			C	00	00	0000	
		0I	00	0000	24			00	00	000I	27
	C	00	00	0000		I2330	C	00	00	0000	
		00	40	0000				00	00	0000	
	C	00	00	0000			C	40	00	0000	
		00	20	0000				00	00	0000	
	C	00	00	0000			C	20	00	0000	
		00	I0	0000				00	00	0000	
	C	00	00	0000			C	I0	00	0000	
		00	04	0000				00	00	0000	
	C	00	00	0000			C	04	00	0000	
		00	02	0000				00	00	0000	
	C	00	00	0000			C	02	00	0000	
		00	0I	0000	25			00	00	0000	28
I23I4	C	00	00	0000			C	0I	00	0000	
		00	00	4000				00	00	0000	
	C	00	00	0000			C	00	40	0000	
		00	00	2000				00	00	0000	
	C	00	00	0000		I2340	C	00	20	0000	
		00	00	I000				00	00	0000	
	C	00	00	0000			C	00	I0	0000	
		00	00	0400				00	00	0000	
I2320	C	00	00	0000			C	00	04	0000	
		00	00	0200				00	00	0000	
		00	00	0000			C	00	02	0000	

		00 00 0000	29		00 00 0000	32
I2344	C	00 01 0000		C	00 00 0060	
		00 00 0000		C	00 00 0000	
	C	00 00 4000		C	00 00 0000	
		00 00 0000		II	010 0367	
	C	00 00 2000	I2370	K	II 013 0366	
		00 00 0000		II	000 0367	
	C	00 00 1000		K	II 000 0074	
		00 00 0000		00	30I 2103	
I2350	C	00 00 0400		K	00 000 0000	
		00 00 0000		00	000 0000	
	C	00 00 0200		K	00 000 0000	
		00 00 0000	30	II	000 007I	33
	C	00 00 0100	I2374	K	II 010 046I	
		00 00 0000		00	30I 2166	
	C	00 00 0040		K	00 000 0000	
		00 00 0000		07	013 0000	
	C	00 00 0020		K	II 000 046I	
		00 00 0000		00	30I 2176	
	C	00 00 0010		K	00 000 0000	
		00 00 0000		00	000 0020	34
	C	00 00 0004		B	00 000 0000	
		00 00 0000		K	00 30I 2246	
	C	00 00 0002		00	000 0000	
		00 00 0000	31	00	000 1000	35
I2360	C	00 00 000I		B	00 000 0000	
		77 77 7777		00	00 0000	
	C	77 77 7777		C	00 00 0000	
		00 00 0000		00	00I 6500	36T
	C	00 00 0100		BC	00 000 0000	
		40 50 0000		I4	24I 2462	
	C	00 00 0000		K	00 000 0000C	
		64 00 0000		I4	010 0000	
	C	00 00 0000		K	00 040 000I	
		64 00 0000		I4	010 000I	
	C	00 00 0100		K	00 26I 2127	

	00	040	0002	
K	I4	250	000I	
	00	30I	2I07	
K	00	000	0000	
	I4	0I0	0000	37T
K	00	040	0002	
	0I	250	000I	
K	0I	0I0	0000	
	02	0I2	0000	
K	00	26I	650I	
	00	30I	2I07	
K	00	000	0000	
	00	042	000I	
K	IJ	000	0070	
	00	042	0002	
K	II	000	007I	
	00	30I	2246	38T
K	00	000	0000	

ПРОГРАММА П4

"Диагностика объектов голосованием по тестам" для ЭВМ БЭСМ-6 составлена в машинных кодах.

Слуцкая Т.Л.

Назначение.

Программа П4 производит классификацию объектов, используя процедуру "голосование по тестам", описанную в [7].

Описание программы.

Так как программа П4 требует для своей работы просмотра всех тупиковых тестов, она включает в себя программу П2.

Две перфокарты I и II (так называемые п/к связи) обеспечивают связь с программой П2 и настройку ее на режим работы в составе П4.

Так как основные затраты машинного времени по-прежнему приходится на перебор тестов, оказывается выгодным обрабатывать сразу группу объектов $\{S_j\}$. Программа П4 допускает одновременную классификацию m' объектов экзамена; $m' \leq 380$.

Ограничения на размеры таблицы Т, в частности, на число столбцов-признаков n , те же, что и в П2 ($n \leq 48$, $m \leq 1024$).

Блок-схема программы П4

Программа П4 распадается на 4 основных блока.

Блок I (яч. 00100-00107) - начальные засылки, очистка массива результатов, расшифровка информации о прочерках в объектах S_j ;

Блок 2 (I2070-I2377) - соответствующая настроенная программа П2;

Блок 3 (00110-00137) - блок подсчета "голосов", обращение к блоку 3 из блока 2 происходит после получения очередного тупикового теста;

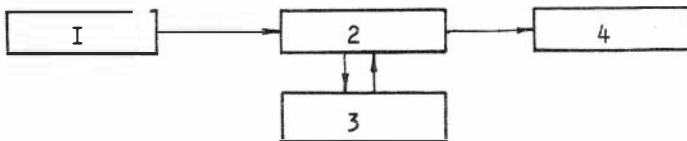
для каждого исследуемого объекта S_j просматривается вся таблица Т (т.к. при наличии прочерков у S_j возможен случай, когда один тест голосует за несколько эталонов) и к соответствующим счетчикам прибавляется по единице;

в начале блока производится "упрятивание" регистров, т.е. заполнение их содержимого, а в конце - восстановление;

Блок 4 (00140-00205) - работает один раз в конце работы программы;

обрабатываются и печатаются результаты в естественном порядке, а также по убыванию.

Блок-схема программы выглядит следующим образом:



Распределение памяти

00100-00205 - блоки 1,3,4 программы П4;

α_1 - 00577 - рабочие ячейки; начало переменное. $\alpha_1 = 00577 - m_{(8)}$

где $m_{(8)}$ - число строк в таблице T в восьмеричной системе; $\alpha_1 \geq 00206$.

00600- α_2 - массив рабочих ячеек, используемых для упорядочения "голосов" по убыванию; $\alpha_2 < 00600 + 2 \cdot m_{(8)}$;

02600- α_3 - счетчики "голосов" для всех эталонов и всех объектов γ ; $\alpha_3 < 12000$;

01200- α_4 } массивы информации, задаваемые пользователем

02000- α_5 } $\alpha_4 < 02000$; $\alpha_5 < 02600$

Инструкция к пользованию

1. Информация о таблице $T_{(m \times n)}$ задается так же, как и для программы П2.

2. Информация о классифицируемых объектах $\{S_j\}$ кодируется восьмеричными цифрами по одному объекту в ячейке, начиная с ячейки 01200; заканчивается массив $\{S_j\}$ нулевым кодом.

3. Если объекты $\{S_j\}$ имеют прочерки, то информация о прочерках объекта γ_2 помещается в ячейку с адресом $02000 + (2 - I)_8$ в виде единиц в разрядах, соответствующих прочеркам, и нулей - во всех остальных разрядах.

Заканчивается массив информации о прочерках также нулевым кодом. Для сокращения длины вводимой информации рекомендуется все объекты S_j , имеющие прочерки, вынести в начало массива.

4. Порядок перфокарт в программе П4 следующий:

- а) паспорт
- б) время;
- в) E ;
- г) 1,2,...,35,(П2);
- д) I,П,...,XV, Z ;
- е) информация к задаче;
- ж) E - конец;

Печать результатов

В программе П4 имеется промежуточная печать через каждые 1024 тупиковых теста так же, как и в П2.

В конце программы печатаются:

а) K - общее число тупиковых тестов;

P_i - информационные веса;

б) $\delta_j^{abc} = K_j/K$, где K_j - число "голосов", полученных j-ым эталоном. $\{\delta_j^{abc}\}$ - печатаются отдельными массивами (через 2 интервала) для всех исследуемых объектов S_j ;

в) для удобства обработки числа δ_j^{abc} для каждого S_j упорядочиваются по убыванию и печатаются под значениями: первое число δ_j^{abc} по убыванию, второе - номера эталонов, которым они соответствуют.

Контрольный пример

№ призн. объект	1	2	3	4	5	6	7	8	9
1	1	0	-	0	-	1	0	0	1
2	0	1	0	1	1	0	1	1	0
3	-	0	1	-	0	0	1	0	1

П

γ^1	1	0	-	1	1	1	-	-	0
γ^2	-	0	0	1	-	1	0	0	0

Т проб

Закодированная информация имеет вид:

B	00	001	2500
	00	000	0000
C	41	10	0000
	00	00	0000
C	26	60	0000
	00	00	0000

Информация о таблице П

C	10	50	0000	Информация о прочер- ках в TI	V	00	000	1200	Информация о классифи- цируемых объектах { δ_j } (пробах)
	00	00	0000		V	00	000	0000	
	00	00	0000		C	47	00	0000	
C	00	00	0000		C	00	00	0000	
	00	000	1000		C	05	00	0000	
V	00	000	0000		C	00	00	0000	
	50	00	0000		C	00	000	0000	
C	00	00	0000		C	00	000	0000	
	00	00	0000		V	00	000	2000	
C	00	00	0000		V	00	000	0000	
	00	000	1030	C	10	60	0000	Информация о прочерках в пробах	
V	00	000	0000	C	00	00	0000		
	12	00	0000	C	42	00	0000		
C	00	00	0000	C	00	00	0000		
	44	00	0000	C	00	00	0000		
C	00	00	0000	C	00	00	0000		
	00	00	0000	C	00	00	0000		
C	00	00	0000						
	00	00	0000						
C	00	00	0000						

Результаты счета сведены в таблицы:

признака	1	2	3	4	5	6	7	8	9
P_i	0	0,2	0,2	0,0	0,2	0,5	0,5	0,2	0,2

№ объекта	δ абс r_1	δ абс r_2
1	0,800	0,800
2	0,400	0
3	0,300	0

ТЕКСТ ПРОГРАММЫ П4

Паспорт :

Шифр xxxxxx -

Лист 1 0-II, 34-37 -

Вход 1 00100 -

Авост -

Время х хх хх -
 час мин сек

	BC	00	001	2070					
		00	000	0000					
		01	247	7717	I		11	000	0076 3
I2070	K	11	241	2000		K	00	271	2226
		00	010	0000			01	241	2500
	K	11	000	0074		K	02	241	2500
		00	221	2461			03	240	0000
	K	01	000	0000		K	07	241	2000
		01	371	2072			01	010	0000
	K	12	240	0167		K	02	012	0001
		01	247	6001			01	011	2000
	K	11	010	0361	I2110	K	02	011	2001
		00	221	6477			11	000	0070
	K	01	000	0000		K	00	261	2123
		01	371	2075	2		03	341	2121 4
	K	02	240	0000		K	03	010	0000
		04	204	0000			11	011	0070
	K	10	311	2103		K	11	000	0071
		02	010	1030			03	012	0000
I2100	K	11	012	0361		K	00	261	2123
		07	000	2177			11	010	0071
	K	02	250	0001		K	11	012	0070
		00	301	2226			00	271	2120
	K	00	000	0000		K	04	010	0000
		01	250	0001			03	000	0000
	K	01	010	0777		K	04	250	0001

I2I20	K	03	37I	2II2	5		II	000	0070	8
		00	000	0000			K II	0I0	036I	
	K	04	044	0005			0I	I00	7777	
		04	257	7777			K 05	37I	2I30	
	K	II	0I0	0070			04	044	0003	
		04	000	0000			K 00	0I0	0000	
	K	04	044	0003			03	000	0000	
	K	02	250	000I			K 03	37I	2I45	
	K	02	0I0	000I			06	250	000I	
	K	00	27I	2I07			K II	0I0	036I	
	K	0I	250	000I			II	000	0070	
	K	0I	044	0002			K II	0I0	0362	
		0I	0I0	000I	6	I2I50	K 00	040	0005	9
	K	00	27I	2I07			06	044	0002	
		II	0I0	036I			04	227	7777	
	K	II	000	0070			K 02	0I0	0000	
	K	II	0I0	0365			II	0II	0070	
I2I30	K	II	000	007I			K II	000	007I	
	K	04	044	0003			II	0I2	0070	
	K	03	250	000I			K 00	30I	2264	
	K	03	II0	7777			04	227	7777	
	K	II	0II	0070			K 02	0I0	0000	
	K	II	022	0364			II	020	007I	
	K	II	000	0072			K 05	036	0000	
	K	II	005	007I	7		02	II5	7777	IO
	K	00	26I	2I37			K 02	I00	7777	
	K	II	0I0	0072			02	37I	2I54	
	K	II	000	007I			K II	0I0	007I	
	K	03	044	000I			II	022	0362	
	K	04	044	0006		I2I60	K II	000	0362	
	K	03	37I	2I32			06	37I	2I50	
	K	0I	II0	7777			K 04	250	000I	
	K	04	220	0000			04	044	000I	
I2I40	K	05	I00	7777			K 00	0I0	0000	
	K	II	0II	0070			0I	II5	7777	
	K	II	0I2	0070			K 0I	37I	2I63	

		II 000 0070	II		I2 240 0000	I4
	K	I0 3II 2220			K II 0I0 0075	
		07 0I0 0000			II 000 0073	
	K	00 30I 2374			K 00 30I 2I74	
		II 000 0073			00 30I 2206	
	K	I2 240 0000		I22I0	K 00 000 0000	
		04 044 000I			00 0I0 0000	
	K	00 000 0000			K II 000 0074	
		0I IIO 7777			I2 240 0005	
I2I70	K	II 0II 0073			K II 0I0 0075	
		I2 26I 2I73			II 000 0076	
	K	0I 37I 2I70			K 00 30I 2252	
		I2 30I 2200	I2		00 000 0000	I5
	K	00 000 0000			K I0 3II 2226	
		0I IIO 7777			II 0I0 0363	
	K	00 000 0000			K 07 237 7500	
		II 000 0070			II 004 0400	
	K	I0 3II 2220			K 07 237 7500	
		II 0I0 0073			II 000 0400	
	K	00 30I 2376			K 00 30I 22I4	
		II 000 0073			00 000 0000	
	K	II 0I2 0360		I2220	K 00 0I0 0000	
		00 27I 2I67			II 000 0074	
	K	00 30I 2240			K II 0I0 0070	
		II 0I0 0073	I3		II 023 0074	I6
I2200	K	II 000 0075			K II 000 0074	
		II 000 0076			00 03I 0000	
	K	00 0I0 0000			K 00 27I 2222	
		II 000 0074			II 230 0074	
	K	00 000 0000			K 07 24I 2300	
		I2 240 00I0			I0 300 0000	
	K	I0 3II 2226			K 00 000 0000	
		II 0I0 0075			II 0I0 0076	
	K	07 0I2 0000			K I2 26I 220I	
		II 000 0073			II 023 0074	
	K	00 30I 2I67			K II 000 0074	

		00 03I 0000	I7			II 010 0400	20
I2230	K	II 000 0076			K	II 005 0260	
		II 230 0074				00 27I 2233	
	K	07 24I 2300			K	II 010 0260	
		IO 300 0000				II 004 026I	
	K	00 000 0000			K	II 000 0260	
		07 240 0300				II 064 0250	
	K	00 30I 22I5			K	II 064 0236	
		00 000 0000				00 000 0000	
	K	00 000 0000			K	00 000 0000	
		00 000 0000				00 30I 2233	
	K	00 000 0000			K	00 000 0000	
		00 00I 2400	I8			45 50 0000	2I
	K	00 00I 246I		I2260	C	00 00 0000	
		03 00I 00I6				45 50 0000	
	K	II 022 0006			C	00 00 0000	
		0I 247 7720				00 00 0000	
I2240	K	00 000 0000			C	00 00 0000	
		00 22I 246I				00 00 0000	
	K	0I 010 0000			C	00 00 0000	
		II 016 0400				II 000 0070	
	K	00 22I 246I			K	00 010 0000	
		0I 000 0000				II 000 0067	
	K	0I 37I 224I			K	II 010 007I	
		II 064 0236	I9			00 26I 2154	22
	K	00 000 0000			K	II 023 0067	
		00 074 0000				II 000 0067	
	K	00 000 0000			K	07 000 000I	
		II 064 0250				00 03I 0000	
	K	00 000 0000		I2270	K	07 37I 2266	
		00 30I 2244				00 000 0000	
	K	00 000 0000			K	00 000 0000	
		00 00I 2070				00 000 0000	
I2250	K	00 00I 2100			K	00 000 0000	
		02 00I 0020				00 000 0000	
	K	IO 023 0004			K	00 000 0000	

		00 00I 230I	23			00 00 0100	26
B		00 00I 210I			C	00 00 0000	
		40 00 0000				00 00 0040	
C		00 00 0000			C	00 00 0000	
		20 00 0000				00 00 0020	
C		00 00 0000			C	00 00 0000	
		10 00 0000				00 00 0010	
C		00 00 0000			C	00 00 0000	
		04 00 0000				00 00 0004	
C		00 00 0000			C	00 00 0000	
		02 00 0000				00 00 0002	
C		00 00 0000			C	00 00 0000	
		01 00 0000	24			00 00 0001	27
C		00 00 0000		I2330	C	00 00 0000	
		00 40 0000				00 00 0000	
C		00 00 0000			C	40 00 0000	
		00 20 0000				00 00 0000	
C		00 00 0000			C	20 00 0000	
		00 10 0000				00 00 0000	
C		00 00 0000			C	10 00 0000	
		00 04 0000				00 00 0000	
C		00 00 0000			C	04 00 0000	
		00 02 0000				00 00 0000	
C		00 00 0000			C	02 00 0000	
		00 01 0000	25			00 00 0000	28
I2314	C	00 00 0000			C	01 00 0000	
		00 00 4000				00 00 0000	
C		00 00 0000			C	00 40 0000	
		00 00 2000				00 00 0000	
C		00 00 0000			C	00 20 0000	
		00 00 1000				00 00 0000	
C		00 00 0000			C	00 10 0000	
		00 00 0400				00 00 0000	
320	C	00 00 0000			C	00 04 0000	
		00 00 0200				00 00 0000	
C		00 00 0000			C	00 02 0000	

I2344	C	00 00 0000	29		C	00 00 0000	32
		00 0I 0000				00 00 0060	
	C	00 00 0000			C	00 00 0000	
		00 00 4000				00 00 0000	
	C	00 00 0000				II 0I0 0367	
		00 00 2000		I2370	K	II 0I3 0366	
	C	00 00 0000				II 000 0367	
		00 00 I000			K	II 000 0074	
50	C	00 00 0000				00 30I 2I03	
		00 00 0400			K	00 000 0000	
	C	00 00 0000				00 000 0000	
		00 00 0200			K	00 000 0000	
	C	00 00 0000	30			II 000 007I	33
		00 00 0I00		I2374	K	II 0I0 046I	
	C	00 00 0000				00 30I 2I66	
		00 00 0040			K	00 000 0000	
	C	00 00 0000				07 0I3 0000	
		00 00 0020			K	II 000 046I	
	C	00 00 0000				00 30I 2I76	
		00 00 00I0			K	00 000 0000	
	C	00 00 0000				00 000 0020	34
		00 00 0004			B	00 000 0000	
	C	00 00 0000				00 30I 2246	
		00 00 0002			K	00 000 0000	
	C	00 00 0000	3I			00 000 I000	35
I2360	C	00 00 000I			B	00 000 0000	
		77 77 7777				00 00 0000	
	C	77 77 7777			C	00 00 0000	
		00 00 0000				п/к связи	
	C	00 00 0I00				00 00I 22I7	I
		40 50 0000			B	00 000 0000	
	C	00 00 0000				II 000 0400	
		64 00 0000			K	00 300 0I32	
	C	00 00 0000				00 00I 2233	
		64 00 0000			B	00 000 0000	
	C	00 00 0I00				07 240 0300	
					K	00 0I0 0000	

		02	000	2600	IX			04	044	0002	XII
	K	04	000	2577		00I60	K	07	045	0004	
		03	00I	00I0				00	064	0I36	
	K	I2	022	0000			K	00	000	0000	
00I40		0I	240	000I				03	370	0I60	
	K	07	240	0000			K	00	000	0000	
		00	22I	2500				02	240	0000	
	K	07	0I0	0000			K	I0	044	00I2	
		00	260	0I44				07	240	060I	
	K	0I	257	7777			K	00	000	0000	
		07	250	000I				02	044	0004	
	K	00	300	0I4I			K	0I	044	0003	
		02	247	7200	X			00	0I0	0000	XIII
	K	I0	240	000I			K	07	000	0000	
		02	0I0	2000				04	0I0	2600	
	K	00	260	0I50			K	07	005	0000	
		I0	257	7777				00	270	0I73	
	K	02	370	0I45		00I70	K	04	0I0	2600	
		00	074	0000				07	000	0000	
	K	00	000	0000			K	03	0I0	0577	
		0I	044	0003				07	000	000I	
00I50	K	02	247	0600			K	04	044	0006	
		00	22I	2000				04	250	000I	
	K	02	0I0	0000			K	03	370	0I67	
		II	0I6	0400	XI			00	0I0	0000	XIV
	K	00	22I	2000			K	06	000	2600	
		02	000	0000				07	250	0002	
	K	02	370	0I5I			K	07	IIO	7776	
		II	0I0	0363				00	270	0I65	
	K	03	000	0577			K	07	I00	7777	
		03	0I0	0577				04	044	0002	
	K	II	004	0363			K	00	000	0000	
		03	000	0600				00	064	0204	
	K	03	370	0I55		00200	K	00	000	0000	
		04	240	0000				I2	370	0I64	
	K	I0	044	0003			K	00	000	0000	

	00	074	0000	XY		00	000	2000	SI(z)
K	00	000	0000		B	00	000	0000	
	00	000	0000			00	00	0000	
K	00	000	0000		C	00	00	0000	
	00	000	060I						
K	07	007	7776						
	03	00I	00I0						
K	I2	022	0006						

ПРОГРАММА П5

"Диагностика объектов голосованием по тесторам"
составлена в машинных кодах для ЭВМ типа
БЭСМ-6.

Слуцкая Т.Л.

Назначение.

Программа П5 производит классификацию объектов, используя процедуру "голосование по тесторам". Кроме того, программа П5 позволяет производить голосование по Q-тестам [3] и пакетам [6].

Инструкция к пользованию

1. Информация о таблицах T_1, T_2, \dots, T_k задается точно так же, как к программе П3, а информация об объектах $\{S_j\}$ — так же, как для П4. То есть:

- а) все таблицы T_1, T_2, \dots, T_k кодируются подряд, как единая таблица и закодированная информация помещается в память, начиная с ячейки I2500. Заканчивается последняя таблица нулевым кодом.
- б) задается информация о началах таблиц T_1, T_2, \dots, T_k в k -ячейках, начиная с I2642, так что в разрядах $i-15$ ячейки $I2642+(i-1)$ стоит код $(I2500 + \sum_{z=1}^{i-1} m_z)$, где m_z — длина T_z ; в ячейку $I2642+k$ помещается нулевой код;
- в) информация о классифицируемых объектах $\{S_j\}$ кодируется восьмеричными цифрами по одному объекту в ячейке, начиная с ячейки OI200; заканчивается массив $\{S_j\}$ нулевым кодом.
- г) Если объекты $\{S_j\}$ имеют прочерки, то информация о прочерках объекта j_2 помещается в ячейку с адресом $O2000+(2-i)_8$ в виде единиц в разрядах, соответствующих прочеркам, и нулей — во всех остальных разрядах. Заканчивается массив информации о прочерках нулевым кодом.

2. Порядок перфокарт в П5.

- 1). Паспорт
- 2). Время
- 3) E
- 4). Программа (G-5I)
- 5). Информация к задаче (а, б, в, г).
- 6). E —конец.

3. Результаты выдаются в следующем порядке:

а) K^* - общее число тупиковых тесторов и P_i^* информационные веса, как в ПЗ;

б) $\delta_j^{*abc}(T)$ - по всем эталонам таблицы T_T , затем их сумма $\sum_{j=1}^{m_i} \delta_j^{*abc}(T) = \delta_{(1)}^{*abc}$ и общее приведенное (поделенное на K^*) число голосов за первую таблицу $\delta^{*abc}(1)^*$

После этого (через 2 интервала) печатаются δ_j^{*abc} и $\delta_{(2)}^{*abc} = \sum_{j=1}^{m_i} \delta_j^{*abc}$, соответственно для второй таблицы и т.д., $\delta_j^{*abc}(T)$, $\delta_{(T)}^{*abc}$, для объекта S_T ($T=I$). Аналогично печатаются результаты для всех $\{S_T\}$ по порядку ($T=I, 2, \dots, m'$).

Программа П5 позволяет вычислить пакетные величины A_i , $\bar{A}_i, B(s), \bar{B}(s)$. Для вычисления A_i и $B(s)$ в качестве Т1 берется Т, а в качестве Т2 - нулевая строка. Тогда $A_i = P_i^*$, а $B(s) = I - \delta_{(2)}^{*abc}$. Для вычисления \bar{A}_i и $\bar{B}(s)$ в качестве Т1 берется Т, а в качестве Т2 - единичная строка. Тогда $\bar{A}_i = P_i^*$, а $\bar{B}(s) = I - \delta_{(2)}^{*abc}$.

Примечание: При подготовке информации добавить к обеим таблицам в I-м случае единичный, а во втором - нулевой отождествляющий столбец для выполнения одного из условий данной программы, предусматривающего отсутствие строки, состоящей из сплошных I или O.

Контрольный пример к П5

ПРИЗН. ОБЪ.										
	I	2	3	4	5	6	7	8	9	
TI	1	0	0	I	I	0	I	I	-	0
II кл.	2	-	I	0	0	0	I	0	I	I
I кл.	3	I	0	0	I	I	0	0	0	I

Результаты счёта

δ_{T1}^{*abc}	δ_{T2}^{*abc}
0,400	0,200
0,100	0
0,200	0,200

За I-й класс

$\Sigma = 0,700 = 0,400$

№ ПР. ОБЪ.										
	I	2	3	4	5	6	7	8	9	
T2	1	-	I	0	0	I	0	-	I	I
II кл.	2	I	0	0	I	0	I	0	0	0

Результаты счёта

δ_{T1}^{*abc}	δ_{T2}^{*abc}
0,100	0
0,600	I,000

За II-й класс

$\Sigma = 0,700 = I,000$

№ ПР. ОБЪ.										
	I	2	3	4	5	6	7	8	9	
T3	1	I	I	I	0	0	0	0	I	0
III кл.	2	0	-	0	I	-	I	I	0	I

Результаты счёта

δ_{T1}^{*abc}	δ_{T2}^{*abc}
0,000	0
0,250	0,200

За III класс

$\Sigma = 0,250 = 0,200$

*) $\delta_j^{*abc}(T)$ - число голосов за j-й эталон T_T , поделенное на K^* .

Т проб	№ пр. объ	1	2	3	4	5	6	7	8	9
	1	1	0	-	1	1	1	-	-	0
2	-	0	0	1	-	1	0	0	0	

Результаты счета:

№ пр.	1	2	3	4	5	6	7	8	9
P_i^*	0,400	0,100	0,550	0,450	0,400	0,600	0,500	0,450	0,550

Закодированная информация имеет вид:

	00	001	2500
B	00	000	0000
	15	40	0000
C	00	00	0000
	21	30	0000
C	00	00	0000
	46	10	0000
C	00	00	0000
	22	30	0000
C	00	00	0000
	45	00	0000
C	00	00	0000
	70	20	0000
C	00	00	0000
	05	50	0000
C	00	00	0000
	00	00	0000
C	00	00	0000
	00	000	1000
B	00	000	0000
	64	40	0000
C	00	00	0000
	00	00	0000
C	00	00	0000
	00	000	1030
B	00	000	0000

Информация о таблицах
Т1, Т2, Т3

Информация о прочерках в
Т1, Т2, Т3.

	00	20	0000	}
С	00	00	0000	
	40	00	0000	
С	00	00	0000	
	40	40	0000	
С	00	00	0000	
	22	00	0000	
С	00	00	0000	
	00	00	0000	
С	00	00	0000	
	00	00I	2462	}
В	00	000	0000	
	00	00	0000	
С	00	0I	2500 T ₁	
	00	00	0000	
С	00	0I	2503 T ₂	
	00	00	0000	
С	00	0I	2505 T ₃	
	00	00	0000	
С	00	00	0000	
	00	000	I200	}
В	00	000	0000	
	47	00	0000	
С	00	00	0000	
	05	00	0000	
С	00	00	0000	
	00	00	0000	
С	00	00	0000	
	00	000	2000	
В	00	000	0000	
	10	60	0000	}
С	00	00	0000	
	42	00	0000	
С	00	00	0000	
	00	00	0000	
С	00	00	0000	

Информация о началах таблиц

Информация о классифицируемых объектах (пробах) $\{S_r\}$

Информация о прочерках в пробах

Результаты счета приведены в таблицах контрольного примера.

ТЕКСТ ПРОГРАММЫ П5 В МАШИННЫХ КОДАХ

Паспорт :

Шифр \square xxxxxx

Лист \square 0-II, 34-37

Вход \square 00100

Авост

Время \square х хх хх
час мин сек

	В	00	001	2070							
		00	000	0000							
I2070	К	01	247	7717	I	К	11	000	0076	3	
		11	241	2000			00	271	2226		
		00	010	0000			03	240	0000		
	К	11	000	0074			К	07	241	2000	
		00	221	2461				00	301	6500	
	К	01	000	0000				К	00	000	0000
		01	371	2072				01	010	0000	
	К	12	240	0167				К	02	012	0001
		01	247	6001				01	011	2000	
	К	11	010	0361	I2110		К	02	011	2000	
		00	221	6477				11	000	0070	
	К	01	000	0000				К	00	261	6511
		01	371	2075	2			03	341	2121	4
	К	02	240	0000				К	03	010	0000
		04	240	0000					11	011	0070
	К	10	311	2103				К	11	000	0071
		02	010	1030					03	012	0000
I2100	К	11	012	0361				К	00	261	2123
		07	000	2177					11	010	0071
	К	02	250	0001				К	11	012	0070
		00	301	2226					00	271	2120
	К	00	000	0000				К	04	010	0000
		01	250	0001					03	000	0000
	К	01	010	0777				К	04	250	0001

		03	37I	2II2	5		II	000	0070	8
I2I20	K	00	000	0000			K	II	0I0	036I
		04	044	0005				0I	I00	7777
	K	04	257	7777			K	05	37I	2I30
		II	0I0	0070				04	044	0003
	K	04	000	0000			K	00	0I0	0000
		04	044	0003				03	000	0000
	K	02	250	000I			K	03	37I	2I45
		02	0I0	000I				06	250	000I
	K	00	27I	2I07			K	II	0I0	036I
		00	30I	6505				II	000	0070
	K	00	000	0000			K	II	0I0	0362
		0I	0I0	000I	6			00	040	0005
	K	00	27I	2I07		I2I50	K	06	044	0002
		II	0I0	036I				04	227	7777
	K	II	000	0070			K	02	0I0	0000
I2I30	K	II	0I0	0365				II	0II	0070
		II	000	007I			K	II	000	007I
	K	04	044	0003				II	0I2	0070
		03	250	000I			K	00	30I	2264
	K	03	IIO	7777				04	227	7777
		II	0II	0070			K	02	0I0	0000
	K	II	022	0364				II	020	007I
		II	000	0072			K	05	036	0000
	K	II	005	007I	7			02	II5	7777
		00	26I	2I37			K	02	I00	7777
	K	II	0I0	0072				02	37I	2I54
		II	000	007I			K	II	0I0	007I
	K	03	044	000I				II	022	0362
		04	044	0006		I2I60	K	II	000	0362
	K	03	37I	2I32				06	37I	2I50
		0I	IIO	7777			K	04	250	000I
I2I40	K	04	220	0000				04	044	000I
		05	I00	7777			K	00	0I0	0000
	K	II	0II	0070				0I	II5	7777
		II	0I2	0070			K	0I	37I	2I63

	II	000	0070	II		I2	240	0000	I4
K	IO	3II	2220			K	II	0IO	0075
	007	0IO	0000				II	000	0073
K	00	3OI	2374			K	00	3OI	2174
	II	000	0073				00	3OI	2206
K	I2	240	0000		I22IO	K	00	000	0000
	04	044	000I				00	0IO	0000
K	00	000	0000			K	II	000	0074
	0I	IIO	7777				I2	240	0005
I2I70	K	II	0II	0073		K	II	0IO	0075
	I2	26I	2I73				II	000	0076
K	0I	37I	2I70			K	00	3OI	2252
	I2	3OI	2200	I2			00	000	0000
K	00	000	0000			K	IO	3II	2226
	0I	IIO	7777				II	0IO	0363
K	00	000	0000			K	07	237	7500
	II	000	0070				II	004	0400
K	IO	3II	2220			K	07	237	7500
	II	0IO	0073				II	000	0400
K	00	3OI	2376			K	00	3OI	22I4
	II	000	0073				00	000	0000
K	II	0I2	0360		I2220	K	00	0IO	0000
	00	27I	2I67				II	000	0074
K	00	3OI	2240			K	II	0IO	0070
	II	0IO	0073	I3			II	023	0074
I2200	K	II	000	0075		K	II	000	0074
	II	000	0076				00	03I	0000
K	00	0IO	0000			K	00	27I	2222
	II	000	0074				II	230	0074
K	00	000	0000			K	07	24I	2300
	I2	240	00IO				IO	300	0000
K	IO	3II	2226			K	00	000	0000
	II	0IO	0075				II	0IO	0076
K	07	0I2	0000			K	I2	26I	220I
	II	000	0073				II	023	0074
K	00	3OI	2I67			K	II	000	0074

		00	03I	0000	I7		II	010	0400	I9
I2230	K	II	000	0076		K	II	005	0260	
		II	230	0074			00	27I	2233	
	K	07	24I	2300		K	II	010	0260	
		IO	300	0000			II	004	026I	
	K	00	000	0000		K	II	000	0260	
		07	240	0300			II	064	0250	
	K	00	30I	22I5		K	00	000	0000	
		00	000	0000			II	064	0236	
	K	00	000	0000		K	00	000	0000	
		00	000	0000			00	30I	2233	
	K	00	000	0000		K	00	000	0000	
00		00	00I	2400	I8		00	00I	2264	20
	K	00	00I	246I		B	00	000	0000	
		03	00I	00I6			II	000	0070	
I2240	K	II	022	0006		K	00	010	0000	
		0I	247	7720			II	000	0067	
I2240	K	00	000	0000		K	II	010	007I	
		00	22I	246I			00	26I	2154	
	K	0I	010	0000		K	II	023	0067	
		II	0I6	0400			II	000	0067	
	K	00	22I	246I		K	07	000	000I	
		0I	000	0000			00	03I	0000	
	K	0I	37I	224I		K	07	37I	2266	
		II	064	0236	I9		00	00I	230I	2I
	K	00	000	0000		B	00	00I	210I	
		00	074	0000			40	00	0000	
	K	00	000	0000		C	00	00	0000	
		II	064	0250			20	00	0000	
	K	00	000	0000		C	00	00	0000	
		00	30I	2244			IO	00	0000	
	K	00	000	0000		C	00	00	0000	
		00	00I	2070			04	00	0000	
I2250	K	00	00I	2100		C	00	00	0000	
		02	00I	0020			02	00	0000	
	K	IO	023	0004		C	00	00	0000	

		01 00 0000	22			00 00 000I	25
	C	00 00 0000		I2330	C	00 00 0000	
		00 40 0000				00 00 0000	
	C	00 00 0000			C	40 00 0000	
		00 20 0000				00 00 0000	
	C	00 00 0000			C	20 00 0000	
		00 10 0000				00 00 0000	
	C	00 00 0000			C	10 00 0000	
		00 04 0000				00 00 0000	
	C	00 00 0000			C	04 00 0000	
		00 02 0000				00 00 0000	
	C	00 00 0000			C	02 00 0000	
		00 01 0000	23			00 00 0000	26
I23I4	C	00 00 0000			C	01 00 0000	
		00 00 4000				00 00 0000	
	C	00 00 0000			C	00 40 0000	
		00 00 2000				00 00 0000	
	C	00 00 0000		I2340	C	00 20 0000	
		00 00 1000				00 00 0000	
	C	00 00 0000			C	00 10 0000	
		00 00 0400				00 00 0000	
320	C	00 00 0000			C	00 04 0000	
		00 00 0200				00 00 0000	
	C	00 00 0000			C	00 02 0000	
		00 00 0100	24			00 00 0000	27
	C	00 00 0000		I2344	C	00 01 0000	
		00 00 0040				00 00 0000	
	C	00 00 0000			C	00 00 4000	
		00 00 0020				00 00 0000	
	C	00 00 0000			C	00 00 2000	
		00 00 0010				00 00 0000	
	C	00 00 0000			C	00 00 1000	
		00 00 0004				00 00 0000	
	C	00 00 0000		50	C	00 00 0400	
		00 00 0002				00 00 0000	
	C	00 00 0000			C	00 00 0200	

	C	00	00	0000	28		B	00	00I	22I3	32
		00	00	0I00				00	000	0000	
		00	00	0000			K	II	000	0076	
	C	00	00	0040				00	30I	2252	
		00	00	0000			B	00	00I	2260	
	C	00	00	0020				00	000	0000	
		00	00	0000			C	45	50	0000	
	C	00	00	00I0				00	00	0000	
		00	00	0000				45	50	0000	
	C	00	00	0004			C	00	00	0000	
		00	00	0000				00	00	0000	
	C	00	00	0000			C	00	00	0000	
		00	00	0002				00	00	0000	
		00	00	0000	29		B	00	000	I000	33
I2360	C	00	00	000I				00	000	0000	
		77	77	7777			C	00	00	0000	
	C	77	77	7777				00	00	0000	
		00	00	0000			B	00	00I	2374	34
	C	00	00	0I00				00	000	0000	
		40	50	0000			K	II	000	007I	35
	C	00	00	0000		I2374		II	0I0	046I	
		64	00	0000			K	00	30I	2I66	
	C	00	00	0000				00	000	0000	
		64	00	0000			K	07	0I3	0000	
	C	00	00	0I00				II	000	046I	
		00	00	0000	30		K	00	30I	2I76	
	C	00	00	0060				00	000	0000	
		00	00	0000			B	00	00I	6500	36T
	C	00	00	0000				00	000	0000	
		II	0I0	0367			K	I4	24I	2462	
I2370	K	II	0I3	0366				00	000	0000	
		II	000	0367			K	I4	0I0	0000	
	K	II	000	0074				00	040	000I	
		00	30I	2I03			K	I4	0I0	000I	
	K	00	000	0000				00	26I	2I27	
		00	000	0020	3I		K	00	040	0002	
	B	00	000	0000				I4	250	000I	
		00	30I	2246			K	00	30I	2I07	
	K	00	000	0000				00	000	0000	

K	I4	010	0000	37		01	247	0600	IY	
	00	040	0002		0100	K	00	010	0000	
	01	250	0001			K	00	221	2000	
K	01	010	0000			K	01	000	0000	
	02	012	0000			K	01	370	0101	
K	00	261	6501			K	01	247	7201	
	00	301	2107			K	01	010	2577	
K	00	000	0000			K	00	260	0106	
	00	042	0001			K	00	012	0131	
K	11	000	0070			K	01	000	2577	
	00	042	0002			K	01	370	0103	
K	11	000	0071			K	00	301	2070	
	00	301	2246	38		K	00	010	0131	Y
K	00	000	0000			K	01	000	2577	
	00	001	2217	I		K	01	370	0106	
B	00	000	0000		0110	K	00	301	2070	
	11	000	0400			K	00	000	0000	
K	00	300	0132			K	I3	247	7770	
	00	001	2233			K	I3	042	0011	
B	00	000	0000			K	I3	000	0050	
	07	240	0300			K	I3	370	0111	
K	00	010	0000			K	03	240	0000	
	00	000	0027			K	00	000	0000	
K	11	000	0300			K	01	240	0000	
	00	301	2215			K	00	010	0027	YI
K	00	000	0000			K	01	011	2000	
	00	001	2212	II		K	00	000	0030	
B	00	000	0000			K	02	241	2500	
	I2	246	5707			K	02	010	0000	
K	11	010	0075			K	00	260	0125	
	00	001	2245			K	01	012	I200	
B	00	000	0000		0120	K	02	011	2000	
	00	300	0140			K	00	011	0030	
K	00	000	0000			K	00	260	0123	
	00	000	0100	III		K	03	250	0001	
B	00	000	0000			K	02	250	0001	

00I22	K	00	300	01I6	УП	00	040	0002	Xa		
		00	000	0000		K	05	250		000I	
		03	010	2600			00	010		0000	
	K	II	004	0363		K	00	000	0077		
		03	000	2600			03	044	0004		
	K	00	300	0I2I		K	00	000	0000		
		0I	250	000I			03	010	2600		
	K	0I	010	I200		K	II	0I6	0400		
		00	270	0II4			03	000	2600		
	K	I3	247	7770	00I50	K	00	004	0077		
I3		010	0050			00	000	0077			
I3		040	00II	K		03	250	000I			
00I30	K	I3	370	0I27	УШ	K	0I	250	000I	XIa	
		00	30I	2206			0I	010	0000		
	K	77	77	7777		K	02	0I2	0000		
		77	77	7777			00	270	0I47		
	K	07	237	7500		K	00	064	0I36		
		II	010	0300			00	000	0000		
	K	00	0I3	0027		K	00	064	0I62		
		00	000	0027			00	000	0000		
	K	00	30I	22I4		K	00	300	0I42		
		00	000	0000			00	000	0000		
K	00	000	0000			06	010	I200			
	00	000	0000		K	06	250	000I			
00I36	K	04	000	2600	IXa	I60	K	00	270	0I4I	XIa
		03	000	2577				00	000	0000	
	K	03	00I	00I0		K	00	074	0000		
		II	022	0000			00	000	0000		
00I40	K	06	240	000I		K	00	000	0077		
		03	240	0000			00	000	0077		
	K	05	24I	2462		K	03	005	00II		
		00	000	0000			I3	022	0000		
	K	05	010	0000		B	00	000	2000	5I	
		00	260	0I57			00	000	0000		
K	00	040	000I		C	00	00	0000			
	05	010	000I			00	00	0000			

ПРОГРАММА П6

"Вычисление тестовых и тесторных параметров таблиц с учетом длин тупиковых тестов и тесторов". Программа написана в машинных кодах для ЭВМ типа БЭСМ-6.

Слущкая Т.Л.

Назначение.

Программа П6 для заданной группы таблиц вычисляет число k всех тупиковых тесторов, числа K_i - тупиковых тесторов длины l , включающих i -ый столбец и ряд других информационных характеристик столбцов и строк.

Если каждую строку данной таблицы T принять за одну из таблиц набора, то по указанной программе можно получить аналогичные характеристики распределения длин тупиковых тестов для таблицы T .

Инструкция к пользованию.

Программа П6 включает в себя как составную часть программу П3. Перфокарта 53 осуществляет настройку программы П3 для работы в режиме П6 и связь с блоком, реализующим вычисление и печать перечисленных ниже величин. Дополнительными являются перфокарты 36-56.

Подготовка исходной информации.

1. Исходная информация кодируется в виде восьмеричных чисел и помещается в память начиная с ячейки I2500; информация о прочерках помещается в ячейки OIO30, ей сопутствующие шкалы, начиная с OIOOO. Все таблицы T_1, T_2, \dots, T_k кодируются одним массивом, заканчивается информация нулевым кодом.

2. Задается информация о началах таблиц T_1, T_2, \dots, T_k в ячейках, начиная с ячейки I2462, так что в разрядах I-I5 ячейки I2462+ ($i-1$) стоит код $(I2500 + \sum_{z=1}^{i-1} m_z)$, где m_z - длина T_z , в ячейку I2642+k помещается нулевой код.

Ограничения.

Общая длина таблиц не должна превышать IO20 слов. В случае, если в разных таблицах окажутся одинаковые строки, произойдет останов и в двух первых напечатанных словах (в разрядах I-I5) будут стоять адреса совпадающих строк.

Порядок постановки перфокарт

1. Паспорт
2. Время
3. Е
4. Программа (0-56)
5. Информация
6. Е - конец

На печать выдаются:

- а) значения $J(s)$, $R(s)$ и $O(s)$ для всех строк - для тестов.
- б) K^l для $l=1,2,\dots,48$; k, l, \bar{L} ;
- в) для каждого i -го признака ($1 \leq i \leq n$): K_i^l для $l=1,2,\dots,48$;
 $\bar{l}_i, P_i, \rho_i, R_i, O_i$.

Контрольный пример

Имеем три таблицы T1, T2, T3, каждая из которых состоит из одной строки и 9 столбцов.

<u>№ пр</u> <u>№ таб.</u>	1	2	3	4	5	6	7	8	9
T1	1	0	0	0	0	1	0	0	1
T2	0	1	0	1	1	0	1	1	0
T3	0	0	1	0	0	0	1	0	1

Закодированная информация имеет вид:

	00	001	2500
B	00	000	0000
	41	10	0000
C	00	00	0000
	26	60	0000
C	00	00	0000
	10	50	0000
C	00	00	0000
	00	00	0000
C	00	00	0000

Информация о таблицах
(T1, T2, T3)

B	00	00I	2462	}	Информация о началах таблиц	
	00	000	0000			
	00	000	0000			
C	00	0I	2500			T ₁
	00	00	0000			
C	00	0I	250I			T ₂
	00	00	0000			
C	00	0I	2502			T ₃
	00	00	0000			
C	00	00	0000			

Результаты счета разносятся в таблицы.

№	1	2	3	4	5	6	7	8	9
\bar{I}_i	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000
P_i	0,26I	0,174	0,348	0,174	0,174	0,26I	0,26I	0,174	0,174
ρ_i	0	0	0	0	0	0	0	0	0
R_i	0	0	0	0	0	0	0	0	0
O_i	0,26I	0,174	0,348	0,174	0,174	0,26I	0,26I	0,174	0,174
K_i^e	6	4	8	4	4	6	6	4	4

Результаты счета:

№	J (s)	R (s)	O (s)
1	0,695	0	0,695
2	0,956	0	0,956
3	0,782	0	0,782

$$K_{\ell=2}^e = 23; K = 23; \bar{L} = 2$$

Паспорт :
 Шифр \sqsubset xxxxxx -
 Лист \sqsubset 0-I, 5-16, 34-37 -
 Вход \sqsubset I6500 -
 Авост -
 Время \sqsubset х хх хх -
 час мин сек

		00	00I	2070							
	B	00	000	0000							
		0I	247	77I7	I		II	000	0076	3	
I2070	K	II	24I	2000			K	00	27I	2226	
		00	0I0	0000				0I	24I	2500	
	K	II	000	0074			K	02	24I	2500	
		00	22I	246I				03	240	0000	
	K	0I	000	0000			K	07	24I	2000	
		0I	37I	2072				0I	0I0	0000	
	K	I2	240	0I67			K	02	0I2	000I	
		0I	247	600I				0I	0II	2000	
	K	II	0I0	036I	I2II0		K	02	0II	200I	
		00	22I	6477				II	000	0070	
	K	0I	000	0000			K	00	26I	2I23	
		0I	37I	2075	2			03	34I	2I2I	4
	K	02	240	0000			K	03	0I0	0000	
		04	240	0000				II	0II	0070	
	K	I0	3II	2I03			K	II	000	007I	
		02	0I0	I030				03	0I2	0000	
I2I00	K	II	0I2	036I			K	00	26I	2I23	
		07	000	2I77				II	0I0	007I	
	K	02	250	000I			K	II	0I2	0070	
		00	30I	2226				00	27I	2I20	
	K	00	000	0000			K	04	0I0	0000	
		0I	250	000I				03	000	0000	
	K	0I	0I0	0777			K	04	250	000I	

		03	37I	2II2	5		II	000	0070	8	
I2I20	K	00	000	0000			K	II	0I0	036I	
		04	044	0005				0I	I00	7777	
	K	04	257	7777			K	05	37I	2I30	
		II	0I0	0070				04	044	0003	
	K	04	000	0000			K	00	0I0	0000	
		04	044	0003				03	000	0000	
	K	02	250	000I			K	03	37I	2I45	
		02	0I0	000I				06	250	000I	
	K	00	27I	2I07			K	II	0I0	036I	
		0I	250	000I				II	000	0070	
	K	0I	044	0002			K	II	0I0	0362	
		0I	0I0	000I	6			00	040	0005	9
	K	00	27I	2I07		I2I50	K	06	044	0002	
		II	0I0	036I				04	227	7777	
	K	II	000	0070			K	02	0I0	0000	
		II	0I0	0365				II	0II	0070	
I2I30	K	II	000	007I			K	II	000	007I	
		04	044	0003				II	0I2	0070	
	K	03	250	000I			K	00	30I	2264	
		03	II0	7777				04	227	7777	
	K	II	0II	0070			K	02	0I0	0000	
		II	022	0364				II	020	007I	
	K	II	000	0072			K	05	036	0000	
		II	005	007I	7			02	II5	7777	IO
	K	00	26I	2I37			K	02	I00	7777	
		II	0I0	0072				02	37I	2I54	
	K	II	000	007I			K	II	0I0	007I	
		03	044	000I				II	022	0362	
	K	04	044	0006		I2I60	K	II	000	0362	
		03	37I	2I32				06	37I	2I50	
	K	0I	II0	7777			K	04	250	000I	
		04	220	0000				04	044	000I	
I2I40	K	05	I00	7777			K	00	0I0	0000	
		II	0II	0070				0I	II5	7777	
	K	II	0I2	0070			K	0I	37I	2I63	

	II	000	0070	II		I2	240	0000	I4
K	I0	3II	2220			K	II	0I0	0075
	07	0I0	0000				II	000	0073
K	00	30I	2374			K	00	30I	2I74
	II	000	0073				00	30I	2206
K	I2	240	0000		I22I0	K	00	000	0000
	04	044	000I				00	0I0	0000
K	00	000	0000			K	II	000	0074
	0I	IIO	7777				I2	240	0005
I2I70	K	II	0II	0073		K	II	0I0	0075
	I2	26I	2I73				II	000	0076
K	0I	37I	2I70			K	00	30I	2252
	I2	30I	2200	I2			00	000	0000
K	00	000	0000			K	I0	3II	2226
	0I	IIO	7777				II	0I0	0363
K	00	000	0000			K	07	237	7500
	II	000	0070				II	004	0400
K	I0	3II	2220			K	07	237	7500
	II	0I0	0073				II	000	0400
K	00	30I	2376			K	00	30I	22I4
	II	000	0073				00	000	0000
K	II	0I2	0360		I2220	K	00	0I0	0000
	00	27I	2I67				II	000	0074
K	00	30I	2240			K	II	0I0	0070
	II	0I0	0073	I3			II	023	0074
I2200	K	II	000	0075		K	II	000	0074
	II	000	0076				00	03I	0000
K	00	0I0	0000			K	00	27I	2222
	II	000	0074				II	230	0074
K	00	000	0000			K	07	24I	2300
	I2	240	00I0				I0	300	0000
K	I0	3II	2226			K	00	000	0000
	II	0I0	0075				II	0I0	0076
K	07	0I2	0000			K	I2	26I	220I
	II	000	0073				II	023	0074
K	00	30I	2I67			K	II	000	0074

		00	03I	0000	I7		II	010	0400	20	
I2230	K	II	000	0076			K	II	005	0260	
		II	230	0074				00	27I	2233	
	K	07	24I	2300			K	II	010	0260	
		IO	300	0000				II	004	026I	
	K	00	000	0000			K	II	000	0260	
		07	240	0300				II	064	0250	
	K	00	30I	22I5			K	00	000	0000	
		00	000	0000				II	064	0236	
	K	00	000	0000			K	00	000	0000	
		00	000	0000				00	30I	2233	
	K	00	000	0000			K	00	000	0000	
		00	00I	2400	I8	I2260	C	45	50	0000	2I
	K	00	00I	246I				00	00	0000	
		03	00I	00I6			C	45	50	0000	
	K	II	022	0006				00	00	0000	
I2240	K	0I	247	7720			C	00	00	0000	
		00	000	0000				00	00	0000	
	K	00	22I	246I			C	00	00	0000	
		0I	0IO	0000				00	00	0000	
	K	II	0I6	0400			K	II	000	0070	
		00	22I	246I				00	0IO	0000	
	K	0I	000	0000			K	II	000	0067	
		0I	37I	224I				II	0IO	007I	
	K	II	064	0236	I9		K	00	26I	2I54	22
		00	000	0000				II	023	0067	
	K	II	074	0000			K	II	000	0067	
		00	000	0000				07	000	000I	
	K	II	064	0250		I2270	K	00	03I	0000	
		00	000	0000				07	37I	2266	
	K	00	30I	2244			K	00	000	0000	
		00	000	0000				00	000	0000	
	K	00	00I	2070			K	00	000	0000	
I2250	K	00	00I	2I00				00	000	0000	
		02	00I	0020			K	00	000	0000	
	K	IO	023	0004				00	000	0000	

	B	00	00I	230I	23		C	00	00	0I00	26
		00	00I	2I0I			C	00	00	0000	
	C	40	00	0000			C	00	00	0040	
		00	00	0000			C	00	00	0000	
	C	20	00	0000			C	00	00	0020	
		00	00	0000			C	00	00	0000	
	C	I0	00	0000			C	00	00	00I0	
		00	00	0000			C	00	00	0000	
	C	04	00	0000			C	00	00	0004	
		00	00	0000			C	00	00	0000	
	C	02	00	0000			C	00	00	0002	
		00	00	0000			C	00	00	0000	
	C	0I	00	0000	24		C	00	00	000I	27
		00	00	0000		I2330	C	00	00	0000	
	C	00	40	0000			C	00	00	0000	
		00	00	0000			C	40	00	0000	
	C	00	20	0000			C	00	00	0000	
		00	00	0000			C	20	00	0000	
	C	00	I0	0000			C	00	00	0000	
		00	00	0000			C	I0	00	0000	
	C	00	04	0000			C	00	00	0000	
		00	00	0000			C	04	00	0000	
	C	00	02	0000			C	00	00	0000	
		00	00	0000			C	02	00	0000	
I23I4	C	00	0I	0000	25		C	00	00	0000	28
		00	00	0000			C	0I	00	0000	
	C	00	00	4000			C	00	00	0000	
		00	00	0000			C	00	40	0000	
	C	00	00	2000			C	00	00	0000	
		00	00	0000		I2340	C	00	20	0000	
	C	00	00	I000			C	00	00	0000	
		00	00	0000			C	00	I0	0000	
	C	00	00	0400			C	00	00	0000	
320		00	00	0000			C	00	04	0000	
	C	00	00	0200			C	00	00	0000	
		00	00	0000			C	00	02	0000	

		00	00	0000	29		00	00	0000	32	
I2344	C	00	01	0000		C	00	00	0060		
		00	00	0000			00	00	0000		
	C	00	00	4000		C	00	00	0000		
		00	00	0000			II	010	0367		
	C	00	00	2000		I2370	K	II	013	0366	
		00	00	0000				II	000	0367	
	C	00	00	1000			K	II	000	0074	
		00	00	0000				00	30I	2103	
50	C	00	00	0400			K	00	000	0000	
		00	00	0000				00	000	0000	
	C	00	00	0200			K	00	000	0000	
		00	00	0000	30			II	000	007I	33
	C	00	00	0100		I2374	K	II	010	046I	
		00	00	0000				00	30I	2166	
	C	00	00	0040			K	00	000	0000	
		00	00	0000				07	013	0000	
	C	00	00	0020			K	II	000	046I	
		00	00	0000				00	30I	2176	
	C	00	00	0010			K	00	000	0000	
		00	00	0000				00	000	0020	34
	C	00	00	0004			B	00	000	0000	
		00	00	0000				00	30I	2246	
	C	00	00	0002			K	00	000	0000	
		00	00	0000	31			00	000	1000	35
I2360	C	00	00	0000			B	00	000	0000	
		00	00	0001				00	00	0000	
	C	77	77	7777			C	00	00	0000	
		77	77	7777				00	00I	6500	36
	C	00	00	0000			B	00	000	0000	
		00	00	0100				00	000	0000	
	C	40	50	0000							
		00	00	0000							
	C	64	00	0000							
		00	00	0000							
	C	64	00	0000							
		00	00	0100							

K	I3	246	3000	37
	00	010	0000	
K	00	223	4000	
	I3	0000	0000	
K	I3	37I	650I	
	II	24I	2000	
K	I3	247	772I	
	0I	242	0000	
K	00	22I	7057	
	I3	010	0000	
K	II	004	0363	
	00	22I	7060	
K	I3	000	0000	38
	00	042	000I	
K	00	22I	7140	
	I3	000	0000	
K	0I	250	0100	
	I3	37I	6504	
K	00	042	00II	
	00	000	0100	
K	00	010	0000	
	00	000	0102	
K	00	000	0103	
	00	30I	2070	
K	II	010	0075	39
	00	022	0100	
K	00	040	00I3	
	I3	010	5000	
K	00	004	0103	
	00	000	0103	
K	I3	010	5000	
	00	000	0104	
K	00	005	0102	
	00	27I	6522	
K	I3	010	5000	
	00	000	0102	

	II	010	0363	40
K	I3	004	5700	
	I3	000	5700	
K	00	30I	22I5	
	00	000	0000	
K	00	000	0000	
	07	237	7500	
K	II	010	5060	
	00	000	010I	
K	00	040	00I3	
	00	010	0104	
K	I3	004	006I	
	I3	000	006I	4I
K	II	010	0075	
	00	022	010I	
K	00	040	00I3	
	II	010	0363	
K	I3	004	0000	
	I3	000	0000	
K	00	30I	22I4	
	00	000	0000	
K	00	000	0000	
	00	000	0000	
K	00	000	0000	
	I3	247	772I	42
K	0I	242	0000	
	II	220	0000	
K	I3	010	0460	
	0I	000	0062	
K	00	26I	6550	
	0I	010	006I	
K	0I	016	0062	
	0I	000	006I	
K	00	016	0102	
	II	006	0363	
K	0I	000	0063	

K	01	010	0062	43
	11	016	0400	
	01	000	0062	
K	01	017	0063	
	01	000	0064	
K	01	006	0062	
	01	000	0065	
K	00	000	0000	
	01	250	0100	
K	I3	37I	6537	
	I3	24I	2000	
K	00	000	0000	
	00	000	0000	44
	11	010	0363	
K	00	000	0101	
	I2	240	4365	
K	00	000	0000	
	I3	010	0500	
K	00	26I	6574	
	11	000	0076	
K	00	010	0000	
	11	000	0074	
K	10	311	2226	
	07	230	4560	45
K	00	010	0064	
	00	017	0101	
K	01	004	2000	
	01	000	2000	
K	07	230	4560	
	00	010	0062	
K	00	017	0101	
	01	004	0000	
K	01	000	0000	
	00	30I	2226	
K	00	000	0000	

K	I2	240	4370	46
	11	010	4612	
	00	000	0101	
K	I3	010	2500	
	11	012	0361	
K	00	271	6556	
	01	010	0000	
K	01	005	2000	
	01	000	4000	
K	01	250	0001	
	I3	250	0001	
K	00	30I	6553	
	00	010	0103	47
K	11	016	0400	
	11	000	5762	
K	11	010	0400	
	11	000	5761	
K	04	247	7721	
	02	247	7776	
K	03	242	6000	
	11	064	4614	
K	00	000	0000	
	01	250	2000	
K	03	250	2000	
	02	371	6600	48
K	01	241	7700	
	01	010	0062	
K	00	261	6622	
	11	064	4616	
K	00	000	0000	
	11	064	4620	
K	00	000	0000	
	01	250	0100	
K	04	371	6603	
	00	0074	0000	
K	00	000	0000	

	00	00I	66I3	49
K	00	00I	66I3	
	03	00I	00II	
K	I3	0I5	00II	
	40	I0	0000	
C	00	00	0000	
	00	000	0000	
K	00	000	0000	
	03	000	0000	
K	0I	007	7777	
	03	00I	00II	
K	I2	0I5	00II	
	0I	000	000I	50
K	0I	000	0060	
	03	00I	00II	
K	I2	0I5	00II	
	0I	000	006I	
K	0I	000	0065	
	03	00I	00II	
K	I3	0I5	00II	
	II	064	46I0	
K	00	000	0000	
	00	30I	6606	
K	00	000	0000	
	II	0I0	007I	5I
K	II	023	4627	
	II	0I2	036I	
K	00	040	0004	
	00	30I	6577	
K	00	000	0000	
	77	77	7777	
C	77	77	7776	
	00	00I	6576	
B	00	000	0000	
	II	000	576I	
K	00	30I	6624	
	00	00I	2075	56
B	00	000	0000	
	00	22I	6447	
K	0I	000	0000	

	00	00I	22I7	52
B	00	000	0000	
	II	000	0400	
K	00	30I	6525	
	00	00I	2240	
B	00	000	0000	
	00	30I	6536	
K	00	000	0000	
	00	00I	2233	
B	00	000	0000	
	07	240	0300	
K	00	30I	65I4	
	00	00I	6450	53
B	00	000	0000	
	I4	24I	2462	54
K	00	000	0000	
	I4	0I0	0000	
K	00	040	000I	
	I4	0I0	000I	
K	00	26I	2I27	
	00	040	0002	
K	I4	250	000I	
	00	30I	2I07	
K	00	000	0000	
	I4	0I0	0000	
K	00	040	0002	
	0I	250	000I	55
K	0I	0I0	0000	
	02	0I2	0000	
K	00	26I	645I	
	00	30I	2I07	
K	00	000	0000	
	00	042	000I	
K	II	000	0070	
	00	042	0002	
K	II	000	007I	
	00	30I	2246	
K	00	000	0000	

ПРОГРАММА П7 "СОСТАВ-1"

составлена в машинных кодах для ЭВМ типа М-220, М-222.

Смертин Е.А.

Назначение

Программа предназначена для вычисления информационных весов столбцов (P_i и Q_i) и строк ($J(S)$ и $Q(S)$) для 5-строчных таблиц с произвольным числом столбцов.*)

Инструкция к пользованию

Кодирование информации.

а) В ячейке 0330 по 2-му адресу ставится число таблиц, которые надо просчитать, $q \leq 0012_8$.

Пример I. Пусть надо просчитать 8 таблиц. Тогда в ячейку 0330 следует поставить:

к	КОП	A1	A2	A3
0	00	0000	0010	0000

б) В ячейку $[0330 + I7_8(i-1) + J]$, ($i = 1, 2, \dots, q \leq 12_8$; $j = 1, 2, \dots, I7_8$) ставится тройка чисел $\langle a_{ij}, b_{ij}, \bar{b}_{ij} \rangle$, где $a_{ij} = b_{ij} + \bar{b}_{ij}$, b_{ij} - число столбцов типа j группы I в i -й таблице, \bar{b}_{ij} - число столбцов типа j группы II в i -й таблице (см. таблицы I, 2).

Таблица I

Разбиение столбцов I группы на типы

№ типа	I	2	3	4	5	6	7	10	11	12	13	14	15	16	17
вид столбца	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1
i -й таб-	0	0	0	1	0	0	0	0	1	1	1	0	0	0	1
лицы	0	0	1	0	0	0	1	1	0	0	1	0	0	1	0
	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0
	1	0	0	0	0	1	1	0	1	0	0	1	0	0	0

*) Здесь $Q(S)$ - сумма Q_i , соответствующих единицам строки S .

Таблица 2

Разбиение столбцов II группы на типы

№ типа	I	2	3	4	5	6	7	10	11	12	13	14	15	16	17
вид столбца	I	I	I	I	0	I	I	I	I	I	I	0	0	0	0
i-й таб-	I	I	I	0	I	I	I	I	0	0	0	I	I	I	0
лицы	I	I	0	I	I	I	0	0	I	I	0	I	I	0	I
	I	0	I	I	I	0	I	0	I	0	I	I	0	I	I
	0	I	I	I	I	0	0	I	0	I	I	0	I	I	I

Пример 2. Пусть $q = 10_8$, как в примере I и в 7-й таблице 5 столбцов I-го типа группы I и 6 столбцов I-го типа группы II. Тогда в ячейку (0330+17 x 6 + I), т.е. в ячейку⁰⁴⁶³ надо поставить:

л	КОП	A1	A2	A3
0	0 0	0013	0005	0006

В итоге исходная информация располагается так:

№ ячейки	л	КОП	A1	A2	A3	КА
	-	-	0330	-	-	
0330	-	-	-	$q \leq 0012$	-	информация к I-й таблице
0331	-	-	a_{11}	b_{11}	\bar{b}_{11}	
0332	-	-	a_{12}	b_{12}	\bar{b}_{12}	
...	информация ко 2-й таблице и т.д.
0347	-	-	$a_{1\ 17}$	$b_{1\ 17}$	$\bar{b}_{1\ 17}$	
0350	-	-	$a_{2\ 1}$	$b_{2\ 1}$	$\bar{b}_{2\ 1}$	
0351	-	-	$a_{2\ 2}$	$b_{2\ 2}$	$\bar{b}_{2\ 2}$	
...	
0366	-	-	$a_{2\ 17}$	$b_{2\ 17}$	$\bar{b}_{2\ 17}$	

Перфокарты с информацией ставятся в конце колоды.

Получение результатов. Для каждой из q таблиц выдается: после I теста печати идет: число всех тупиковых тестов в таблице, число тупиковых тестов, в которые входит столбец I-го типа, число тупиковых тестов, в которые входит столбец 2-го типа, ..., число тупиковых тестов, в которые входит столбец I7-го типа; затем идет информационный вес столбца I-го типа, информаци-

онный вес столбца 2-го типа, ..., информационный вес столбца 17-го типа;

далее идет: информационный вес 1-го эталона, информационный вес 2-го эталона, ..., информационный вес 5-го эталона;

после II теста печати выдаются аналогичные Q-тестовые результаты.

Контрольный пример.

№ пр. объ.	1	2	3	4	5	6	7	8	9
1	0	1	1	1	1	0	1	0	0
2	0	0	1	0	0	1	0	1	0
3	1	0	0	0	0	1	1	0	0
4	0	0	1	0	0	0	1	0	0
5	0	0	1	0	0	0	0	0	1

Закодированная информация имеет вид:

к	КОП	A1	A2	A3	
		0330			КА
		0001	0001		
		0001	0001		
0	00				
		0002	0001	0001	
		0001	0001		
		0003	0003		
0	00				
0	00				
0	00				
		0001		0001	
0	00				
		0001	0001		
0	00				
0	00				
0	00				

Для удобства обработки результатов рекомендуется составлять таблицу вида:

Общий вид

№ типа столбца	Перечень номеров столбцов данного типа	число столбцов типа
I	$\alpha_1, \alpha_2, \dots, \alpha_i$	$A_{i1} = l$
2	$\beta_1, \beta_2, \dots, \beta_g$	$A_{i2} = g$
3
4
5
6
7
10
11
12
13
14
15
16
17	$\delta_1, \delta_2, \dots, \delta_h$	$A_{i7} = h$

Для контрольного примера

№ типа столбца	Перечень номеров столбцов данного типа	число столбцов типа	P_i	Q_i
I	9	I	0,700	0,517
2		0	0	0
3	1,3	2	0,400	0,379
4	8	I	0,500	0,483
5	2,4,5	3	0,333	0,276
6		0	0	0
7		0	0	0
10		0	0	0
11	7	I	0,500	0,241
12		0	0	0
13	6	I	0,400	0,172
14		0	0	0
15		0	0	0
16		0	0	0
17		0	0	0

Результаты счета вынесены в таблицы:

№ призна.	1	2	3	4	5	6	7	8	9
P_i	0,400	0,333	0,400	0,333	0,333	0,400	0,500	0,500	0,700
Q_i	0,379	0,276	0,378	0,276	0,276	0,172	0,241	0,483	0,517

№ объекта	$J(s)$	$Q(s)$
I	I,900	I,448
2	I,300	I,034
3	I,300	0,793
4	0,900	0,621
5	I,100	0,896

$K^P = 30$ (число P-тестов)

$K^Q = 29$ (число Q-тестов)

117

ТЕКСТ ПРОГРАММЫ П7 В МАШИНЫХ КОДАХ

№	П/К	0 00	000I	0000	0000	КА	5	0 55	063I	0640	0000	
	I	0 56	0000	0660	0000			0 36	0000	0625	0000	
		0 00	0660	0000	0000	КА		4 0I	000I	0557	0557	
		0 15	0000	0330	0000			0 56	0000	0626	0000	
		0 36	0000	0667	0000			4 0I	0I0I	0557	0557	
		0 33	0330	0670	0330			0 34	0633	063I	063I	
		0 16	0664	0563	0657			I 12	00I6	062I	000I	
		0 13	0603	0672	0603			0 56	0000	0644	0000	
		0 13	0564	067I	0564			0 00	0000	0000	0000	
		0 56	0000	0660	0000			0 00	4000	0000	0000	
		0 77	0000	0000	0000			0 77	020I	7000	0000	
		0 00	0000	000I	0000			0 00	04I6	I000	0000	
2		0 00	00I7	0000	0000		6	0 00	I062	2000	0000	
		0 00	0000	00I7	0000			0 00	2I24	4000	0000	
		0 00	0560	0000	0000	КА		0 00	4I5I	0000	0000	
		0 00	7777	0000	0000			0 00	0000	0000	0000	
	I	I 4	0000	0000	0000			0 52	0000	0000	0000	
	I	0I	4000	0000	0000			4 00	0633	0000	0640	
		0 52	0000	0000	0000			0 56	0000	06I6	0000	
		5 55	033I	0560	0I0I			0 00	0000	0000	0000	
		5 75	0I0I	056I	0I0I			I 00	0557	0000	0I57	
		5 0I	0I0I	0000	0I0I			I 12	0004	0642	000I	
		I 12	00I6	0564	000I			0 50	00I3	0000	7767	
		0 16	057I	0I70	0322			0 70	7500	0647	0000	
3		0 00	0I64	0000	0I20		7	0 16	0652	750I	76I0	
		0 52	0000	0000	0000			0 52	0I20	0027	0I37	
		5 02	0I0I	0562	0I0I			0 16	0654	750I	76I0	
	I	I 36	0000	0577	0I2I			0 52	0I40	0027	0I56	
		0 16	0576	0I70	0322			0 16	0656	750I	76I0	
	I	02	0I20	0I64	0I2I			0 52	0I57	0027	0I63	
		5 04	0I2I	0I20	0I40			0 00	0I70	0000	0000	КА
		5 0I	0I0I	0562	0I0I			0 52	0000	0000	0000	
	I	I 12	00I6	0573	000I			5 00	0I0I	0000	000I	
		0 52	0000	0000	0000			I 12	00I6	0I7I	000I	
	3	34	056I	033I	000I			0 00	0000	0000	0I64	
	3	34	056I	000I	000I			0 16	0I75	0020	0067	
4		5 55	000I	0560	000I		8	0 00	0000	0000	00I6	
		5 75	000I	056I	000I			0 00	0II6	0000	00I7	
		5 0I	000I	0000	000I			0 00	0II2	0000	00I0	
		5 75	0I0I	056I	0I0I			0 00	0III	0000	0007	
		5 0I	0I0I	0000	0I0I			0 00	0II0	0000	00I2	
	7	05	0I40	000I	000I			0 00	0I07	0000	00II	
	7	05	0I40	0I0I	0I0I			0 00	0I04	0000	0003	
	I	I 12	00I6	0603	000I			0 00	0I03	0000	0004	
		0 56	0000	064I	0000			0 16	0206	0020	0067	
		0 00	0632	0000	063I			0 00	0000	0000	00I5	
		0 00	0000	0000	0557			0 00	0II5	0000	00I7	
	4	52	0000	0000	0644			0 00	0II3	0000	00I0	

№ п/к

9	0 00 0112 0000 0013	I3	0 00 0102 0000 0005
	0 00 0107 0000 0006		0 00 0101 0000 0001
	0 00 0106 0000 0011		0 16 0274 0020 0067
	0 00 0103 0000 0002		0 00 0000 0000 0011
	0 00 0102 0000 0004		0 00 0107 0000 0017
	0 16 0217 0020 0067		0 00 0102 0000 0001
	0 00 0000 0000 0014		0 00 0101 0000 0005
	0 00 0114 0000 0017		0 16 0301 0020 0067
	0 00 0112 0000 0007		0 00 0000 0000 0014
	0 00 0111 0000 0013		0 00 0106 0000 0017
	0 00 0110 0000 0006		0 00 0103 0000 0001
	0 00 0107 0000 0012		0 00 0102 0000 0004
I0	0 00 0102 0000 0001	I4	0 16 0306 0020 0067
	0 00 0101 0000 0004		0 01 0102 0101 0070
	0 16 0230 0020 0067		0 05 0103 0070 0070
	0 00 0000 0000 0007		0 05 0104 0070 0070
	0 00 0000 0000 0010		0 05 0105 0070 0070
	0 00 0000 0000 0013		0 01 0070 0164 0164
	0 00 0113 0000 0017		0 01 0104 0105 0070
	0 00 0112 0000 0014		0 05 0103 0070 0070
	0 00 0111 0000 0015		0 05 0104 0105 0071
	0 00 0110 0000 0011		0 01 0071 0070 0070
	0 00 0106 0000 0006		0 05 0102 0070 0070
	0 00 0105 0000 0003		0 05 0101 0070 0070
II	0 00 0104 0000 0005	I5	0 01 0070 0164 0164
	0 00 0103 0000 0004		0 00 0020 0000 0000
	0 00 0101 0000 0002		0 01 0003 0006 0072
	0 16 0245 0020 0067		0 01 0001 0002 0070
	0 00 0000 0000 0014		0 05 0072 0070 0070
	0 00 0112 0000 0017		0 01 0002 0007 0073
	0 00 0107 0000 0006		0 05 0001 0073 0071
	0 00 0106 0000 0012		0 01 0070 0071 0070
	0 00 0103 0000 0001		0 01 0002 0003 0071
	0 00 0102 0000 0004		0 05 0071 0010 0071
	0 16 0254 0020 0067		0 01 0070 0071 0070
	0 00 0000 0000 0015		0 05 0003 0007 0071
I2	0 00 0111 0000 0017	I6	0 01 0070 0071 0070
	0 00 0110 0000 0006		0 01 0004 0005 0071
	0 00 0107 0000 0011		0 05 0070 0071 0070
	0 00 0102 0000 0002		0 01 0072 0073 0075
	0 00 0101 0000 0004		0 05 0011 0075 0071
	0 16 0263 0020 0067		0 01 0070 0071 0070
	0 00 0000 0000 0012		0 01 0012 0013 0071
	0 00 0000 0000 0006		0 01 0071 0075 0071
	0 00 0110 0000 0017		0 05 0071 0014 0071
	0 00 0106 0000 0014		0 01 0070 0071 0070
	0 00 0104 0000 0002		0 01 0001 0010 0074
	0 00 0103 0000 0004		0 01 0074 0011 0074

№	П/К										
I7	0 01	0072	0074	0075	2I	0 05	00I4	I072	I072		
	0 05	00I2	0075	007I		0 0I	I07I	I072	I07I		
	0 0I	0070	007I	0070		0 05	00I5	I07I	I07I		
	0 0I	00I3	00I4	007I		0 0I	I070	I07I	I070		
	0 0I	007I	0075	007I		0 0I	I074	00I3	I07I		
	0 05	00I5	007I	007I		0 0I	I07I	0005	I07I		
	0 0I	0070	007I	0070		0 05	I07I	00I2	I07I		
	0 0I	00I2	0073	0075		0 0I	00I4	I07I	I07I		
	0 0I	0075	0074	0075		0 0I	000I	0003	I074		
	0 05	00I3	0075	007I		0 0I	I074	0005	I072		
	0 0I	0070	007I	0070		0 05	00I3	I072	I072		
	0 0I	00I4	00I5	007I		0 0I	I07I	I072	I07I		
	18	0 0I	007I	0075		007I	22	0 05	0005	00I5	I072
		0 05	00I6	007I		007I		0 05	0006	00I6	I075
0 0I		0070	007I	0070	0 0I	I072		I075	I072		
0 05		0070	00I7	0075	0 05	00I0		I072	I072		
0 0I		0075	0I64	0I64	0 0I	I07I		I072	I07I		
0 00		0604	0000	0000	0 05	00I1		I07I	I07I		
3 34	056I	000I	0I0I	κA	0 0I	I070	I07I	I070			
19	0 00	0570	0000	0000	κA	0 05	0007	00I2	I072		
	0 16	057I	0I70	0323	0 0I	I072	I073	I07I			
	0 00	0575	0000	0000	κA	0 05	00I4	I07I	I07I		
	0 16	0576	0I70	0323	0 05	00I0	I07I	I07I			
	0 00	0I70	0000	0000	κA	0 0I	I070	I07I	I070		
4 52	0000	0000	0322	23	0 05	I072	0006	I07I			
I9 ^a	0 00	0777	0000		0000	κA	0 0I	I074	0004	I072	
	0 05	00I3	00I2		I070	0 05	00I4	I072	I072		
	0 0I	0007	0003		I07I	0 0I	I07I	I072	I07I		
	0 0I	I070	I07I		I072	0 0I	I07I	00I3	I07I		
	0 05	0002	I072		I072	0 05	00I6	I07I	I07I		
	0 0I	0004	I070		I07I	0 0I	I070	I07I	I070		
	0 05	0005	I07I		I07I	0 05	00I7	I070	I075		
	0 0I	I07I	I072		I07I	0 0I	I075	II64	II64		
	0 0I	0006	I070		I072	0 00	0067	0000	0000	κA	
	0 05	0003	I072		I072	0 56	0000	0777	0000		
	0 0I	I07I	I072		I070	0 00	0I70	0000	0000	κA	
	0 0I	0002	0003		I074	24	4 52	0000	0000	I322	
	20	0 0I	I074		00I0		I07I	0 00	0I73	0000	0000
		0 05	000I	I07I	I07I		0 56	0000	II72	0000	
0 0I		I070	I07I	I070	0 00		II72	0000	0000	κA	
0 0I		I074	0004	I074	0 00		0000	0000	0I64		
0 05		00I6	I074	I07I	0 56		0000	0I74	II64		
0 0I		00I2	I07I	I07I	0 00		0I74	0000	0000	κA	
0 05		0006	00I3	I073	0 16		0I75	0020	I067		
0 05		I073	0007	I072	0 00		0205	0000	0000	κA	
0 0I		I07I	I072	I07I	0 16		0206	0020	I067		
0 0I		000I	0002	I074	0 00		02I6	0000	0000	κA	
0 0I		I074	00I6	I072	0 16	02I7	0020	I067			
0 0I	I072	0004	I072								

ПРОГРАММЫ МЕТОДА СУММАРНОГО УЧЕТА МЕР
ПРИУРОЧЕННОСТИ И СОГЛАСОВАНИЯ

В настоящей части опубликованы программы для оценки столбцов и строк таблиц описаний по разработанному В.В.Бабицем, В.О.Красавчиковым, Г.С.Федосеевым методу суммарного учета мер приуроченности и согласования [1], [2] и программы для процедур распознавания и упорядочения геологических объектов, основанных на этих оценках (комплекс алгоритмов и программы "Каскад")

Кроме того, в этот комплекс включаются программы для минимизации признакового пространства как для процедур распознавания, так и для процедур упорядочения и программы оптимального бинарного кодирования признаков.

Данные программы разработаны специально для вышеупомянутого метода. С помощью программ П1 и П2 вычисляются следующие оценки столбцов (оценки информативности признаков).

1. Оценка $\tilde{\varphi}_i$ (П1, П2).

Пусть дана таблица $T = (t_{ki})_{m \times n}$, где $k = 1, \dots, m$, $i = 1, \dots, n$ из "0" и "1", с m строками (объектами) и n столбцами (признаками). Целевой признак x из m вещественных чисел. Строки таблицы упорядочены по убыванию целевого признака. (Здесь $X = Xn \cdot 1$)

Начиная с i -й, каждая строка j_1 из T сравнивается со всеми за ней следующими, причем в результате каждого такого сравнения i -му столбцу ($i = 1, 2, \dots, n$) дается предварительная оценка, равная вычисляемой ниже величине $\tilde{\varphi}_i(j_1, j_2)$.

Пусть сравниваются строки j_1 -я и j_2 -я ($j_1 < j_2$), тогда, принимая x_{j_1} за j_1 -е число массива x , а x_{j_2} - j_2 -е число, имеем:

$\tilde{\varphi}_i(j_1, j_2) = 0$, если по i -му столбцу эти строки совпадают,

$\tilde{\varphi}_i(j_1, j_2) = x_{j_1} - x_{j_2}$, если по i -му столбцу в j_1 -й строке стоит "1", а в j_2 -й - "0"

$\tilde{\varphi}_i(j_1, j_2) = x_{j_2} - x_{j_1}$, если по i -му столбцу в j_1 -й "0", а в j_2 -й - "1"

Для данного j_1 суммируем все такие $\tilde{\varphi}_i(j_1, j_2)$, что $j_1 < j_2$ и делим на величину $(m - j_1)$.

Тогда $\tilde{\varphi}_i(j_1, j_2)_n = \sum_{j_2=j_1+1}^m \tilde{\varphi}_i(j_1, j_2) : (m - j_1)$ $\begin{cases} i = 1, \dots, n \\ j = 1, \dots, m - 1 \end{cases}$

Оценки всех столбцов $\tilde{\varphi}_i$ равны: $\tilde{\varphi}_i = \sum_{j_1=1}^{m-1} \tilde{\varphi}_i(j_1, j_2)_n$, $i = 1, \dots, n$

2. Оценка φ_i° (П1). Таблица Т подразделена на две подтаблицы Т1 и Т2, где первые k строк входят в Т1, а в Т2—следующие $m-k$ строк из Т. k считать заданным исходным данным. Предварительные оценки $\varphi_i^{\circ}(j_1, j_2)$ находятся также, как и $\tilde{\varphi}_i(j_1, j_2)$, причем j_1 -я строка берется из Т1, а j_2 обязательно из Т2.*)

$$\varphi_i^{\circ}(j_1, j_2)_n = \sum_{j_2=k+1}^m \varphi_i^{\circ}(j_1, j_2) : (m-k) \cdot k ; \quad i = 1, \dots, n ; \quad j_1 = 1, \dots, k .$$

Оценки столбцов вычисляются по формуле: $\varphi_i^{\circ} = \sum_{j_1=1}^k \varphi_i^{\circ}(j_1, j_2)_n$, $i = 1, \dots, n$.

3. Оценка φ_i^* (П1). Подразделение таблицы Т на Т1 и Т2 при $k = 1, 2, \dots, m-1$ приводит к оценкам $\varphi_i^{\circ}(1), \varphi_i^{\circ}(2), \dots, \varphi_i^{\circ}(m-1)$ по процедуре 2. Для процедуры 3 эти оценки являются предварительными, и окончательные оценки φ_i^* получаются из последовательных вычислений для $i = 1, \dots, n$; $j_1 = 1, \dots, k$; $k = 1, \dots, m-1$.

$$\varphi_i^*(j_1, j_2)_n = \sum_{j_2=k+1}^m \varphi_i^*(j_1, j_2) : (m-k) \cdot k ; \quad \varphi_i^*(k) = \sum_{j_1=1}^k \varphi_i^*(j_1, j_2)_n ; \quad \varphi_i^* = \sum_{k=1}^{m-1} \varphi_i^*(k) .$$

4. Оценка P_i^{\vee} (П2). Таблица Т такая же, как в п.2. Сначала для $j_1=1, \dots, k$ вычисляются оценки $P_i^{\vee, j_1} = \sum_{s=1}^{m-k} (t_{j_1, i} - t_{k+s, i}) : k(m-k)$.

и затем $P_i^{\vee} = \sum_{j_1=1}^k P_i^{\vee, j_1}$ или $P_i^{\vee} = P_i^{\text{I}} - P_i^{\text{II}}$, где $P_i^{\text{I}}, P_i^{\text{II}}$ — частоты выполнения признака X_i в Т1 и Т2.

5. Оценка $\tilde{P}_i^{\vee} = k(m-k) P_i^{\vee}$ (П2).

Кроме того, П1 вычисляет величины $G_i^{\text{I}}, G_i^{\text{II}}$.

$$1) G_i^{\text{I}} = \sum_{k=1}^{m-1} \sum_{s=1}^{m-k} [(t_{k, i} - t_{k+s, i}) : (m-k)] ; \quad 2) G_i^{\text{II}} = \sum_{k=1}^{m-1} \sum_{s=1}^{m-k} (t_{k, i} - t_{k+s, i})$$

Для оценки нагрузок строк по заданной информативности признаков используется следующий подход (программа П3).

Пусть задана бинарная таблица Т с m строками и n столбцами, включающая три класса объектов (I эталонный класс, II эталонный класс, класс проб и объектов экзамена). Для каждого i -го признака ($i \leq n$) задана информативность P_i , причем $-\infty < P_i < +\infty$. В качестве P_i может выступать любая оценка информативности из алгоритма "Каскад" — $P_i^{\vee}, \varphi_i^{\circ}, \tilde{\varphi}_i^{\circ}, \varphi_i^*, \tilde{\varphi}_i^*, \varphi_i^*$. [1, 2]. Требуется вычислить соответствующую нагрузку строки J (P_i) при заданном виде оценки информативности. Подсчет строчечных нагрузок производится следующим образом.

Столбцы упорядочиваются по убыванию величины $|P_i|$.

Если несколько столбцов обладают одинаковым $|P_i|$, то они вы-

*) При практическом использовании φ_i° величина целевого признака для объектов таблицы Т2 зачастую полагается постоянной (например, среднее арифметическое запасов объектов таблицы Т2).

страиваются по увеличению порядкового номера. Затем в каждом столбце производится замена бинарных символов на соответствующие информативности по следующим правилам:

- а) если $P_i > 0$, то "I" i -го столбца заменяются на $|P_i|$, а "0" - присписывается нулевая информативность;
- б) если $P_i < 0$, то "0" i -го столбца заменяются на $|P_i|$, а "I" - присписывается нулевая информативность;
- в) если $P_i = 0$, то и "I", и "0" i -го столбца заменяются на нули.

После указанных преобразований, путем суммирования можно вычислить нагрузку для каждой строки исходной таблицы, причем эту нагрузку можно рассчитать по любому набору признаков из упорядоченного ряда X_1, \dots, X_l , где $1 \leq l \leq n$.

При наличии в исходной таблице лишь одного эталонного класса, нагрузки строк используются для упорядочивания объектов с той или иной степенью соответствия с целевым признаком. В этом случае для использования описываемой программы таблицу необходимо разбить на три формальных класса.

При наличии двух эталонных классов, строчечные нагрузки, кроме того, являются показателем степени разделения объектов разных классов. Для этого на каждом шаге из минимальной строчечной нагрузки в I классе вычитается максимальная строчечная нагрузка из II класса, полученная величина $\rho^1 = J(P_i)_{\min}^I - J(P_i)_{\max}^{II} (-\infty < \rho^1 < \infty)$ указывает на степень "растяжки" классов и используется при процедуре минимизации признакового пространства.

При наличии кроме эталонных классов объектов экзамена и проб строчечные нагрузки позволяют производить внешний экзамен системы распознавания и распознавание проб.

Программа П4 предназначена для реализации процедур бинарного кодирования и основывается на следующих теоретических предпосылках: пусть задана таблица T из m строк и n столбцов, включающая в себя один или два класса объектов. Признаки (столбцы) заданы как в логической (качественной), так и в порядковой и количественной формах.* Следует закодировать исходную таблицу, т.е. привести ее к бинарному виду, наиболее выгодному (оптимальному) для решения задачи по алгоритму "Каскад". Критерием оптимальности кодирования признака является максимальная по абсолютной величине оценка информативности $|P_i|$. Программа позволяет про-

* и признаки, имеющие шкалу наименований [6] .

изводить кодирование с учетом трех разновидностей оценок информативности — P_i^v , $\hat{\psi}_i$ и ψ_i^* . Оценка P_i^v используется при двухклассовой постановке с бинарно заданным целевым признаком, $\hat{\psi}_i$ — с количественно заданным целевым признаком. Оценка ψ_i^* служит критерием оптимальности при одноклассовой постановке задачи с количественно заданным целевым признаком.

Признаки количественного и качественного ^{и порядкового} типов кодируются по методу "скользящей границы", признаки со ~~списком~~ ~~наименований~~ по методу "перебора", порядковые — также по методу "скользящей границы".

Метод "скользящей границы" заключается в следующем. Все значения признака выстраиваются в ряд, характеризующий собой увеличение проявленности данного свойства на объектах. Варианты кодирования возникают при поочередном присвоении первому значению, двум первым значениям, трем первым значениям и т.д. символа "1", а оставшимся — символа "0". Всего вариантов кодирования таким методом будет:

$$V = \psi - 1,$$

где ψ — число различающихся значений признаков.

При использовании метода "перебора" варианты кодирования возникают при последовательном присвоении каждому сочетанию значений по одному, по два, по три и т.д. символа "1", а оставшимся — символа "0". Поскольку замена в закодированном признаке символа "1" на символ "0" и наоборот не изменяет оценки информативности и не влияет на конечные результаты, количество значимых вариантов кодирования признака этим методом будет:

$$V = C_{\psi}^1 + C_{\psi}^2 + \dots + C_{\psi}^{\frac{\psi}{2}}, \text{ когда } \psi - \text{четное число и}$$

$$V = C_{\psi}^1 + C_{\psi}^2 + \dots + C_{\psi}^{\frac{\psi-1}{2}}, \text{ когда } \psi - \text{нечетное число.}$$

Для каждого полученного тем или иным методом варианта кодирования признака вычисляется требуемая информативность, оптимальным считается вариант, получивший по абсолютной величине оценку. В приведенных ниже формулах величины Z_{cp}^I , Z_{cp}^{II} , $Z_{cp,1}^I$, $Z_{cp,1}^{II}$, P_i^I , P_i^{II} вычисляются с учетом априорно заданных или формально получаемых классов, при этом T_1 и T_2 обозначают априорно заданные классы, T_1^j и T_2^j — формально выделяемые классы, a_{sj} — объект с j -ым порядковым номером. Напомним формулы для P_i^v , $\hat{\psi}_i$, ψ_i^* :
 $P_i^v = (P_i^I - P_i^{II})$ при $T_1 = \{S_1, \dots, S_2\}$ и $T_2 = \{S_{2+1}, \dots, S_m\}$;
 $\hat{\psi}_i = \sum_{k=1}^z \hat{\psi}_k^j$, где $\hat{\psi}_k^j = (Z_{cp}^I - Z_{cp}^{II})(P_i^I - P_i^{II}) + P_i^I(Z_{cp,1}^I - Z_{cp,1}^{II})$ при $T_1^j = \{S_1, \dots, S_j\}$, $j=1, \dots, z$ и $T_2 = \{S_{2+1}, \dots, S_m\}$; $\psi_i^* = \sum_{k=1}^{m-1} \psi_k^{*j}$, где $\psi_k^{*j} = (Z_{cp}^I - Z_{cp}^{II})(P_i^I - P_i^{II}) + P_i^I(Z_{cp,1}^I - Z_{cp,1}^{II}) + P_i^{II}(Z_{cp,1}^{II} - Z_{cp,1}^I)$
 при $T_1^j = \{S_1, \dots, S_j\}$ и $T_2^j = \{S_{j+1}, \dots, S_m\}$, $j = 1, \dots, m$

- $\bar{z}_{ср}^I, \bar{z}_{ср}^II$ – средние значения целевого признака для первого и второго классов соответственно;
- $\bar{z}_{ср,1}^I, \bar{z}_{ср,1}^II$ – средние значения целевого признака для первого и второго классов, подсчитанные для объектов с символом "I" в i -ом признаке ($i=I, \dots, n$);
- P_1^I и P_1^{II} – частота встречаемости значения "I" в i -ом признаке в первом и во втором классах;
- z – количество объектов в I классе;
- n – общее количество объектов в T;
- k – количество границ между формально выделяемыми классами.

Л и т е р а т у р а .

1. В.В.Бабич, В.О.Красавчиков, Г.С.Федосеев. Комплексная оценка информативности признаков при решении прогнозных задач. Сб. Тезисы докладов к третьему Всесоюзному совещанию "Состояние и направление исследований по металлогении траппов". Красноярск, 1974.

2. В.В.Бабич, Г.С.Федосеев. Типизация прогнозно-поисковых задач и некоторые подходы к их решению. Сб. Тезисы докладов к третьему Всесоюзному совещанию "Состояние и направление исследований по металлогении траппов". Красноярск, 1974.

3. А.П.Ершов, Г.И.Кожухин, И.В.Поттосин. Руководство к пользованию системой Альфа. ВЦ СО АН СССР, Новосибирск, 1968.

4. В.М.Курочкин, Д.Б.Подшивалов и др. Система БЭСМ-АЛГОЛ. Методическое руководство по программированию. ВЦ МГУ, М., 1969.

5. А.О.Буда, Т.С.Васючкова, А.А.Грановский, С.Э.Козловский, В.И.Шелехов. Руководство к пользованию системой автоматического программирования Альфа-6, Новосибирск, ВЦ СО АН СССР, 1974.

6. Супес П., Винес Дж. Основы теории измерений. В сб. "Психологические измерения", М., 1967.

* Под частотой встречаемости P_1 понимается отношение $\frac{a}{g}$, где a – количество единиц в классе, g – количество объектов в классе.

ПРОГРАММА ПІ
 " Оценка информативности признаков"
 составлена на языке Альфа для ЭВМ типа
 М-220, М-222.

Кандыба В.Н.

Назначение. Программа предназначена для вычисления оценок информативности признаков (нагрузок столбцов) — $\tilde{\varphi}_i$, φ_i^0 , φ_i^* , оценивающих меру согласования (корреляции) косвенного признака x_i с целевым признаком x_{π_i} .

Инструкция к пользованию.

Порядок постановки перфокарт.

I. α схема (18 п/к), $K \Sigma$

2. aI , $K \Sigma$

3. mI , $K \Sigma$

4. nI , $K \Sigma$

5. zI , $K \Sigma$

6. tI ; $K \Sigma$

7. xI , $K \Sigma$

:

8. mR , $K \Sigma$

9. nR , $K \Sigma$

10. zR , $K \Sigma$

11. tR ; $K \Sigma$

12. xR , $K \Sigma$

где aI — число таблиц, подлежащих решению ($aI=R$)

$m\beta$ — число строк в β -ой таблице

$n\beta$ — число столбцов —"

$z\beta$ — число объектов I кл. —"

$t\beta$ — исходные таблицы

$x\beta$ — целевой признак в β -ой таблице

Каждый ввод прокладывается тремя пустыми перфокартами. Решать можно несколько (k) таблиц одновременно.

Вывод.

1). $\tilde{\varphi}_i(j_1, j_2)_{\text{норм.}}$ — ($m-I$) раз по i чисел;

2). $\tilde{\varphi}_i$;

3). $\varphi_i^0(j_1, j_2)_{\text{норм.}}$ — k — раз по i чисел;

4). $\varphi_i^0(k)$;

5). $\varphi_i^*(k)$ — ($m-I$) раз по i чисел;

6). φ_i^*

Время трансляции 7 минут.

Время решения при $t_1[I:16, I:10]$, $t_2[I:5, I:14]$ ($aI=2$) равно 6 минут.

В программе сменным является выражение 2α :

{ логический массив $t [I: m, I: n]$; где $n=45, 90, 135$ и т.д. в зависимости от количества столбцов в одной строке.

Контрольный пример

№	1	2	3	4	5	6	7	8	9	Σ_{n+1}
1	0	0	1	0	0	0	1	0	0	216
2	0	0	1	0	0	1	0	1	0	160
3	1	0	0	0	0	1	1	0	0	105
4	0	0	1	0	0	0	1	0	0	95
5	0	0	1	0	0	0	0	0	1	40

Закодированная информация имеет вид:

Выражение 2α : { логический массив $t [I: m, I: 45]$;

π	КОП	A1	A2	A3
+	01	100	-	-
+	01	100	-	-

перфокарта $aI, K \Sigma$

π	КОП	A1	A2	A3
+	01	500	-	-
+	01	500	-	-

перфокарта $mI, K \Sigma$

π	КОП	A1	A2	A3
+	01	900	-	-
+	01	900	-	-

перфокарта $nI, K \Sigma$

π	КОП	A1	A2	A3
+	01	100	-	-
+	01	100	-	-

перфокарта $zI, K \Sigma$

π	КОП	A1	A2	A3
I	04	-	-	-
I	I2	-	-	-
4	I4	-	-	-
I	04	-	-	-
I	01	-	-	-

перфокарта tI
(таблица перфорируется в режиме "команда" построчно)

+ $K \Sigma_{t_1}$.

Σ	КОП	A1	A2	A3
+	03	216	-	-
+	03	160	-	-
+	03	105	-	-
+	02	95	-	-
+	02	40	-	-

перфокарта $\times n_{i+1}$.

$+ \kappa \Sigma \Sigma_{n+1}$.

Результаты счета вынесены в таблицу:

κ	1	2	3	4	5	6	7	8	9
\tilde{y}_i	- 8,6	0 + 8,6	0	0	+57,4	+105,0	+66	- 171,5	
y_i^0	-27,7	0 +27,7	0	0	-41,7	+ 58,0	-14,0	-44,0	
y_i^*	-26,6	0 +26,6	0	0	+60	+170,6	+86,8	-25,7	

ТЕКСТ ПРОГРАММЫ П4

1. начало целые $m, n, j, i, r, k, l, a, a1$; ВВОД ($a1$); { для $a:=1, \dots, a1$ цикл
 { ВВОД (m, n, r);
2. { логический массив $t [1:m, 1:n]$;
3. вещественный массив $x [1:m], \varphi, c [1:n]$; ВВОД (t, x) ;
для $i:=1, \dots, n$ цикл $c[i] := 0$;
4. для $j:=1, \dots, m-1$ цикл { для $i:=1, \dots, n$ цикл $\varphi [i] := 0$;
 { для $k:=j+1, \dots, m$ цикл для $i:=1, \dots, n$ цикл
5. { если $(t[j, i] \wedge t[k, i]) \vee (\neg t[j, i] \wedge \neg t[k, i])$ то на M1;
6. если $t[j, i] \wedge \neg t[k, i]$ то $\varphi [i] := \varphi [i] + x [j] - x [k]$
7. иначе $\varphi [i] := \varphi [i] + x [k] - x [j]$; M1: } для $i:=1, \dots, n$ цикл $\varphi [i] := \varphi [i] / (m-j)$; ВВОД (φ);
8. для $i:=1, \dots, n$ цикл $c [i] := c [i] + \varphi [i]$ } ; ВВОД (c);
9. для $i:=1, \dots, n$ цикл $c [i] := 0$;

9. для $j:=1, \dots, r$ цикл { для $i:=1, \dots, n$ цикл $\varphi[i]:=0$; { для $k:=r+1, \dots, m$ цикл
для $i:=1, \dots, n$ цикл
- I0. { если $(t[j, i] \wedge t[k, i]) \vee (\neg t[j, i] \wedge \neg t[k, i])$ то на M2 ;
- I1. если $t[j, i] \wedge \neg t[k, i]$ то $\varphi[i] := \varphi[i] + x[j] - x[k]$
- I2. иначе $\varphi[i] := \varphi[i] + x[k] - x[j]$; M2 : } ;
для $i:=1, \dots, n$ цикл $\varphi[i] := \varphi[i] / ((m-r) \times r)$;
- I3. вывод (φ) ; для $i:=1, \dots, n$ цикл $c[i] := c[i] + \varphi[i]$ } ; вывод (c),
для $i:=1, \dots, n$ цикл $c[i] := 0$;
- I4. для $l:=1, \dots, m-1$ цикл { для $i:=1, \dots, n$ цикл $\varphi[i] := 0$;
{ для $j:=1, \dots, l$ цикл для $k:=l+1, \dots, m$ цикл
для $i:=1, \dots, n$ цикл
- I5. { если $(t[j, i] \wedge t[k, i]) \vee (\neg t[j, i] \wedge \neg t[k, i])$ то на M3 ;
- I6. если $t[j, i] \wedge \neg t[k, i]$ то $\varphi[i] := \varphi[i] + x[j] - x[k]$
- I7. иначе $\varphi[i] := \varphi[i] + x[k] - x[j]$; M3 : } ; для $i:=1, \dots, n$ цикл $\varphi[i] := \varphi[i] / ((m-l) \times l)$;
- I8. вывод (φ) ; для $i:=1, \dots, n$ цикл $c[i] := c[i] + \varphi[i]$ } } ;
вывод (c) } } } ; конец *

ПРОГРАММА П2

"Оценка информативности признаков"
составлена на языке Альфа для ЭВМ
М-220, М-222

Кандыба Е.Н.

Назначение. Программа предназначена для вычисления оценок информативности признаков (нагрузок столбцов) $\varphi_i, \theta_i^1, \theta_i^2$, оценивающих меру согласования (корреляции) косвенного признака x_i с целевым признаком x_{n+1} для одного класса объектов.

Инструкция к пользованию.

Порядок ввода:

$aI, K\Sigma$ - количество таблиц, подлежащих решению.

$m, K\Sigma$ - число строк исходной таблицы.

$n, K\Sigma$ - число столбцов исходной таблицы.

$t, K\Sigma$ - исходная таблица (из "0" и "1").

$x_{n+1}, K\Sigma$ - целевой признак.

Каждая строка $t[j, i]$ кодируется отдельно. В теле программы выражение 2α подлежит изменению:

{ логический массив $t[I:m, I:n^*]$; где $n^*=45 \cdot z$, z - число строк перфокарты, занятых под одну строку $t[j, i]$.

Вывод: φ_i - для всех $i = 1, \dots, n$

θ_i^1 - для всех $i = 1, \dots, n$

θ_i^2 - для всех $i = 1, \dots, n$

Контрольный пример.

Имеется таблица Т, состоящая из 5 строк и 9 столбцов.

№	1	2	3	4	5	6	7	8	9	x_{n+1}
1	0	0	1	0	0	0	1	0	0	216
2	0	0	1	0	0	1	0	1	0	160
3	1	0	0	0	0	1	1	0	0	105
4	0	0	1	0	0	0	1	0	0	95
5	0	0	1	0	0	0	0	0	1	40

Закодированная информация имеет вид:

x	КОП	A1	A2	A3
+	01	100	-	-
+	01	100	-	- KΣ

перфокарта $aI, K\Sigma$

π	КОП	AI	A2	A3
+	01	500	-	-
+	01	500	-	- $\kappa \Sigma$

перфокарта $m, \kappa \Sigma$

π	КОП	AI	A2	A3
+	01	900	-	-
+	01	900	-	- $\kappa \Sigma$

перфокарта $n, \kappa \Sigma$

π	КОП	AI	A2	A3
I	04	-	-	-
I	I2	-	-	-
4	I4	-	-	-
I	04	-	-	-
I	01	-	-	-

перфокарта t

+ $\kappa \Sigma_t$

π	КОП	AI	A2	A3
+	03	216	-	-
+	03	160	-	-
+	03	105	-	-
+	02	95	-	-
+	02	40	-	-

перфокарта x_{n+1}

+ $\kappa \Sigma_{T_{n+1}}$

Результаты счета вынесены в таблицу:

N	1	2	3	4	5	6	7	8	9
ψ_i	- 8,6	0	+8,6	0	0	+57,4	+105,5	+36,0	-171,5
G_i^1	+0,416	0	-0,416	0	0	+1,166	+1,333	+0,75	-2,083
G_i^2	0	0	0	0	0	+2,0	+2,0	+2,0	-4,0

ТЕКСТ ПРОГРАММЫ П2

- I. начало целые $m, n, j, i, y, k, z, a, a_1, b, d$; ввод (a_1);
 { для $a:=1, \dots, a_1$ цикл { ввод (m, n);
2. логический массив $t [1:m, 1:n]$;
3. вещественный массив $x [1:m]$, $\varphi, c, b, d, [1:n]$;
целый массив $z, y, [1:n]$;
ввод (t, x) ;
4. для $i:=1, \dots, n$ цикл { $c[i]:=0$; $d[i]:=0$; $z[i]:=0$ } ;
5. для $j:=1, \dots, m-1$ цикл { для $i:=1, \dots, n$ цикл { $\varphi[i]:=0$; $y[i]:=0$; $b[i]:=0$ } ;
6. { для $k:=j+1, \dots, m$ цикл для $i:=1, \dots, n$ цикл
7. { если $(t[j, i] \wedge t[k, i]) \vee (\neg t[j, i] \wedge \neg t[k, i])$ то на M_1 ;
8. если $t[j, i] \wedge \neg t[k, i]$ то { $\varphi[i] := \varphi[i] + x[j] - x[k]$; $b[i] := b[i] + 1$ }
9. иначе { $\varphi[i] := \varphi[i] + x[k] - x[j]$; $b[i] := b[i] - 1$ } ; M_1 ; }
10. для $i:=1, \dots, n$ цикл { $\varphi[i] := \varphi[i] / (m-j)$; $y[i] := b[i]$; $b[i] := b[i] / (m-j)$ } ;
11. для $i:=1, \dots, n$ цикл { $c[i] := c[i] + \varphi[i]$; $z[i] := z[i] + y[i]$; $d[i] := d[i] + b[i]$ } ;
12. вывод (c, d, z)} ; конец *

ПРОГРАММА ПЗ

"Подсчет строчечных нагрузок по заданной информативности для бинарных таблиц"
составлена на языке Альфа-6 для ЭВМ БЭСМ-6

Бабич В.В.

Назначение. Программа предназначена для пошагового подсчета нагрузок строк по заданной информативности признаков (информативность, в частности, может выражаться отрицательными числами). Кроме того, программа осуществляет пошаговый подсчет межклассовых расстояний для случая двух классов (таблиц), определяет минимальную пошаговую нагрузку в первом классе и максимальную во втором, что позволяет по максимальному расстоянию между классами минимизировать признаковое пространство и производить распознавание проб и объектов экзамена.

Инструкция к пользованию.

Порядок постановки перфокарт.

1. Система Альфа-6 (66 п/к).
2. Сменная перфокарта (п/к 67): ТАБ(n, K_1, K_2, K_3, m) конец
 n – количество столбцов в Т,
 K_1 – количество строк в I классе,
 K_2 – количество строк во II классе,
 K_3 – количество строк в III классе,
 m – общее количество строк в Т.
3. Исходный массив M1 – порядковые номера столбцов.
4. Исходный массив M2 – порядковые номера строк.
5. Исходный массив M3 – I класс.
6. Исходный массив M4 – II класс.
7. Исходный массив M5 – III класс.
8. Информационные веса признаков.

Данная программа позволяет обрабатывать одновременно несколько таблиц T1, T2, ..., Tl. Данные о таблицах располагаются друг за другом в указанном выше порядке. При этом сменная п/к должна иметь вид:

ТАБ ($n^1, x_1^1, x_2^1, x_3^1, m^1$); ТАБ ($n^2, x_1^2, x_2^2, x_3^2, m^2$); ...; ТАБ ($n^l, x_1^l, x_2^l, x_3^l, m^l$) конец

На печать выдается:

1. Распечатка исходных массивов.
2. Порядковые номера столбцов после упорядочивания по $|R_i|$.
3. Пошаговые строчечные нагрузки $J(R_i)$ для объектов I класса.
4. Пошаговые строчечные нагрузки $J(R_i)$ для объектов II класса.
5. Пошаговые строчечные нагрузки $J(R_i)$ для объектов III класса.
6. Пошаговые \min -е строчечные нагрузки $J(P)_{\min}^I$ для объектов I-го класса.
7. Пошаговые \max -е строчечные нагрузки $J(P)_{\max}^I$ для объектов II класса.
8. Пошаговая величина "растяжки" классов ρ^l .
9. Информационные веса признаков в порядке убывания $|R_i|$.

Перфорация исходных данных производится построчно на устройстве КУ-3. Элементы в массивах отделяются запятой, каждая строка пробивается с новой перфокарты. Массивы отделяются друг от друга точкой с запятой, ТАБ - прописные буквы).

При объеме исходной Т [$n=130, m=20$] время решения составляет 1 мин., время трансляции - 25 сек.

Контрольный пример.

I класс

№	I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	1	1	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0
2	1	0	1	1	1	1	0	1	0	0	1	0	0	1	0	1	0	0	1	1
3	1	1	1	0	1	0	0	1	1	0	0	1	0	0	0	0	1	1	1	0

II класс

№	I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	1	0	1	0	0	1	1	0	1	1	1	1	1	0	0	0	1	0	0
2	0	0	1	1	0	0	1	0	0	0	1	1	0	1	1	0	1	1	0	0

III класс

про- ба	I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	1	1	1	0	1	0	0	0	1	0	1	1	1	0	0	0	0	1	0	0

№	I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
ρ_i^I	0,166	0,666	0,500	-0,333	0,666	0,666	0,500	-0,333	+0,666	-0,500	-0,666	-0,666	-0,500	-0,333	-0,500	-0,333	-0,166	-0,666	0,666	0,333

Закодированная информация имеет вид:

I,2,3,4,5,6,7,8,9,I0,II,I2,I3,I4,I5,I6,I7,I8,I9,20; -M1

I,2,3,4,5,6 ; - M2

0,I,I,I,0,I,0,0,I,0,0,0,0,I,0,0,0,0,0,0, } - M3
 I,0,I,I,I,I,0,I,0,0,I,0,0,I,0,I,0,0,I,I,
 I,I,I,0,I,0,0,I,I,0,0,I,0,0,0,0,I,I,I,0;

I,0,0,I,0,0,I,I,0,I,I,I,I,I,0,0,0,I,0,0, } - M4
 0,0,I,I,0,0,0,I,0,0,I,I,0,I,I,0,I,I,0,0;

I,I,I,0,I,0,0,0,I,0,I,I,I,I,0,0,0,0,I,0,0; } - M5

0,I66, 0.666,0.500,-0.333,0.666,0.666,-0.500,-0.333, } P_i
 0.666,-0.500,-0.666,-0.666,-0.500,-0.333,-0.500,
 -0.333,-0.I66,-0.666,0.666,0.333;

ТАБ (20,3,2,I,6) конец - п/к 67

Выдача на печать.

После исходных данных на печать выдаются следующие величины:
 № столбцов, упорядоченных по P_i.

2	5	6	9	II	I2	I8	I9	3	7
I0	I3	I5	4	8	I4	I6	20	I	I7

Пошаговые строчечные нагрузки для объектов I кл. J (P_i)

I	2	3	4	5	6	7	8	9	I0	} объект
0,666	0,666	I,332	I,998	2,664	3,330	3,996	3,996	4,496	4,496	
II	I2	I3	I4	I5	I6	I7	I8	I9	20	} объект
5,496	5,996	6,496	6,496	6,829	6,829	7,162	7,162	7,162	7,328	
I	2	3	4	5	6	7	8	9	I0	} объект
0	0,666	I,332	I,332	I,332	I,998	2,664	3,330	3,830	4,330	
II	I2	I3	I4	I5	I6	I7	I8	I9	20	} объект
4,830	5,330	5,830	6,163	6,163	6,496	6,829	6,829	6,995	6,995	
I	2	3	4	5	6	7	8	9	I0	} объект
0,666	I,332	I,332	I,998	2,664	2,664	2,664	3,330	3,830	4,330	
II	I2	I3	I4	I5	I6	I7	I8	I9	20	} объект
4,830	5,330	5,830	6,163	6,163	6,496	6,829	6,892	6,995	6,995	

Пошаговые строчечные нагрузки для объектов II кл. J (P_i^v)

№	I	2	3	4	5	6	7	8	9	10	} объект	
	P_i^v	0	0	0	0	0	0	0	0	0		
№	II	I2	I3	I4	I5	I6	I7	I8	I9	20		
	P_i^v	0	0	0,500	0,500	0,500	0,500	0,833	0,833	0,999		1,165
№	I	2	3	4	5	6	7	8	9	10		} объект
	P_i^v	0	0	0	0	0	0	0	0,500	1,000		
№	II	I2	I3	I4	I5	I6	I7	I8	I9	20		
	P_i^v	1,500	2,000	2,000	2,000	2,000	2,000	2,333	2,333	2,333		

Пошаговые строчечные нагрузки для объектов III кл. J (P_i^v)

№	I	2	3	4	5	6	7	8	9	10	} объект
	P_i^v	0,666	1,332	1,332	1,998	1,998	1,998	1,998	1,998	2,498	
№	II	I2	I3	I4	I5	I6	I7	I8	I9	20	
	P_i^v	3,498	3,498	3,998	4,331	4,664	4,997	5,330	5,330	5,496	

Пошаговая min-я нагрузка для объектов I-го класса J (P_i^v)_{min}

№	I	2	3	4	5	6	7	8	9	10
	P_i^v	0	0,666	1,332	1,332	1,332	1,998	2,664	3,330	3,830
№	II	I2	I3	I4	I5	I6	I7	I8	I9	20
	P_i^v	4,830	5,330	5,830	5,830	5,830	5,830	5,830	6,163	6,329

Пошаговая max-я строчечная нагрузка для объектов II кл. J (P_i^v)_{max}

№	I	2	3	4	5	6	7	8	9	10
	P_i^v	0	0	0	0	0	0	0	0,500	1,000
№	II	I2	I3	I4	I5	I6	I7	I8	I9	20
	P_i^v	1,500	2,000	2,000	2,000	2,000	2,000	2,333	2,333	2,333

Величина пошаговой "растяжки" классов ρ^1

№	I	2	3	4	5	6	7	8	9	10
	ρ^1	0	0,666	1,332	1,332	1,332	1,998	2,664	3,330	3,330
№	II	I2	I3	I4	I5	I6	I7	I8	I9	20
	ρ^1	3,330	3,330	3,830	3,830	3,830	3,830	3,497	3,830	3,996

Упорядоченные информационные веса признаков (P_i^v)

№	I	2	3	4	5	6	7	8	9	10
P_i^v	0,666	0,666	0,666	0,666	0,666	0,666	0,666	0,666	0,500	0,500
№	11	12	13	14	15	16	17	18	19	20
P_i^v	0,500	0,500	0,500	0,333	0,333	0,333	0,333	0,333	0,166	0,166

ТЕКСТ ПРОГРАММЫ ПЗ

041

```

начало   цел i ; проц таб (n,k1,k2,k3,k) ;
начало цел i,j,q ; цел массив цм6, цм7, цм1[1:n] ,цм2 [1:k], цм3 [1:k1,1:n] ,
цм4 [1:k2,1:n] , цм5 [1:k3, 1:n] ; вещ вп1, вп2 ;
массив вм1 , вм2, вм6, вм7, вм8 [1:n], вм3 [1:k1,1:n] , вм4 [1:k2,1:n] ,
вм5 [1:k3, 1:n] ;
проц печт ( r,k,n);
начало цел i,j ; цел массив цм1 [1:n] ;
для i:=1 шаг 1 до k цикл начало для j:=1 шаг 1 до n
цикл цм1 [j] := r[i,j] ;
OUTPUT (0,'ybd',цм1, '/') конец ; OUTPUT (0,'2/') конец ;
проц пет (r,k,n) ;
начало цел i,j ; массив вм1 [1:n] ;
для i:=1 шаг 1 до k цикл начало для j:=1 шаг 1 до n цикл вм1[j]:=r[i,j] ;
OUTPUT (0,'e'вм1, '/') конец ; OUTPUT (0,'2/') конец ;
проц уп (r,s,k,n) ;
начало цел i,j,q ; для i:=1 шаг 1 до n цикл начало q:= цм7[i];

```

```

для j:=1 шаг 1 до k цикл s [j,i]:=r [j,q] конец конец ;
проц сум (r,k,n) ;
начало цел i,j ; для i:=2 шаг 1 до n цикл
  начало для j:=1 шаг 1 до k цикл r [j,i]:=r [j,i]+ r [j,i -1] конец конец;
проц зам (r,k,n,вм2) ;
начало цел i,j; веш вп 1;
для i:=1 шаг 1 до n цикл начало вп1:=abs (вм2[i] );
если вм2[i]>0 тонам2 ; если вм2[i]<0 то на м3 ;
для j:=1 шаг 1 до k цикл r [j,i] :=0 ; на м4 ; м2 :
для j:=1 шаг 1 до k цикл если r [j,i]>1-10 -8 то r [j,i]:= вп 1; на м4 ;
м3: для j:=1 шаг 1 до k цикл если r [j,i]>1-10-8 то r [j,i] :=0
иначе r [j,i] := вп 1; м4: конец конец ;
INPUT (o,цм 1,цм2,цм3, цм4, цм5, вм 1);
OUTPUT (o,'у4, в4д', цм 1,'2/'); OUTPUT (o,'у4, в4д', цм2,'2/');
печт (цм3,к1, n); печт (цм4,к2, n); печт (цм5,к3, n); OUTPUT (o,'e', вм 1,'2/');
для i:=1 шаг 1 до n цикл начало q:=j:=1 ; если вм1[1]>-5 10 7 то вп 1:=abs (вм1[1])
иначе вп1 :=-10 8; м1: j:=j-1; если вм1[j]>5 10 7 то начало если вп 1 < abs (вм1 [j] ) -10 -8

```

ТО начало $q := j$; $vp1 := abs (vm1[j])$, конец конец ; если $j < n$ то на $m1$;
 $vm2 [i] := vm1 [q]$; $цм7 [i] := q$; $цм6 [i] := цм1 [q]$; $vm1 [q] := -108$ конец ; уп ($цм3, vm3, k1, n$) ;
уп ($цм4, vm4, k2, n$) ; уп ($цм5, vm5, k3, n$) ; зам ($vm3, k1, n, vm2$) ; зам ($vm4, k2, n, vm2$) ;
зам ($vm5, k3, n, vm2$) ; сум ($vm3, k1, n$) ; сум ($vm4, k2, n$) ; сум ($vm5, k3, n$) ;
для $i := 1$ шаг 1 до n цикл начало $vp2 := vm3 [1, i]$;
для $j := 1$ шаг 1 до $k1$ цикл начало $vp1 := vm3 [j, i]$; $vp2 := \min (vp1, vp2)$ конец ;
 $vm6 [i] := vp2$; $vp2 := vm4 [1, i]$;
для $j := 1$ шаг 1 до $k2$ цикл начало $vp1 := vm4 [j, i]$; $vp2 := \max (vp1, vp2)$ конец ;
 $vm7 [i] := vp2$; $vm8 [i] := vm6 [i] - vm7 [i]$ конец ;
OUTPUT ($o, 'y4v4d', цм6, '2/'$) ; OUTPUT ($o, 'y4v4 d' цм2 '2/'$) ;
пет ($vm3, k1, n$) ; пет ($vm4, k2, n$) ; пет ($vm5, k3, n$) ;
OUTPUT ($o, 'e', vm6, '2/'$) ; OUTPUT ($o, 'e', vm7, '2/'$) ;
OUTPUT ($o, 'e', vm8, '2/'$) ; OUTPUT ($o, 'e', vm2, '8/'$) конец ;
таб ($20, 3, 2, 1, 6$) конец

141

Распределение памяти

поле рабочих ячеек	0040I+0055I
поле скаляров	00552+00573
поле констант	00574*006I2
поле рабочей программы	006I3+02065
поле станд. подпрограммы	02066+03620

ПРОГРАММА П4

"Оптимальное бинарное кодирование признаков"
составлена на языке АЛГОЛ-БЭСМ-6 для ЭВМ БЭСМ-6.

Бабич В.В.

Назначение. Программа П4 предназначена для оптимального бинарного кодирования признаков, заданных как в логической (качественной), так и в порядковой и количественной формах. Критерием оптимальности кодирования признака является максимальная по абсолютной величине, определенным образом выбранная оценка информативности. Программа может быть использована как для одноклассовых, так и для двухклассовых постановок задач. (Оценки информативности следующие: $P_i^V, \hat{\psi}_i, \psi_i^*$).

Инструкция к пользованию.

Порядок постановки перфокарт.

1. Схема АЛГОЛ-БЭСМ-6 (68 п/к).
2. Сменная перфокарта п/к 69: ТАБ(n, k_1, k_2, m, d) - END
 n - количество столбцов в T,
 k_1 - количество строк в I классе,
 k_2 - количество строк во II классе,
 m - общее количество строк в T,
 d - указатель оценки информативности, которую следует использовать при кодировании,
при $d=1$ - P_i^V
при $d=2$ - $\hat{\psi}_i$
при $d=3$ - ψ_i^*
3. Исходный массив M1 - порядковые номера столбцов.
4. Исходный массив M2 - массив ключей, указывающих метод, которым необходимо кодировать признак:
I - метод "перебора"
0 - метод "скользящей границы"
5. Исходный массив M3 - значения целевого признака.
6. Исходный массив M4 - I класс
7. Исходный массив M5 - II класс

Данная программа позволяет обрабатывать одновременно несколько таблиц T1, T2, ..., Tl по любой из указанных оценок информативности (несколько задач). Данные о таблицах располагают-

ся друг за другом в указанном выше порядке. При этом сменная перфокарта п/к 69 должна иметь вид:

ТАБ ($n^1, k_1^1, x_2^1, m^1, d^1$);
 ТАБ ($n^2, k_1^2, x_2^2, m^2, d^2$); ... ,
 ТАБ ($n^l, k_1^l, x_2^l, m^l, d^l$) - END ⊐

На печать выдается:

1. Распечатка исходных данных.
2. Исходные таблицы, закодированные оптимальным образом.
3. Оценка информативности признаков.

Перфорация исходных данных производится построчно на устройстве УПП. Элементы в массивах отделяются друг от друга запятой, каждая строка пробивается с новой перфокарты. Массивы отделяются между собой точкой с запятой.

Контрольный пример

№	I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	$T_{л+1}$
1	I	5	7	I8	2	5	I4	4	8	IC1	2	4	3	2	3	I	2	0	I	3	54
2	2	0	4	II2	4	5	IO	II	I	8I	3	3	0	2	I	3	2	I	3	2	38
3	2	4	I5	IO	5	●	4	7	8	I50	0	I	3	I	I	2	I	2	3	I	14
4	2	2	0	23	I	2	I9	7	4	I90	3	I	2	2	I	I	2	3	I	I	4
5	I	I	IO	I2	2	4	IO	6	I	90	I	I	3	2	0	2	I	2	2	3	I

Информация на бланках кодируется следующим образом:

1. Информ. перфокарта (№ 69): ТАБ (20,3,2,5,1); ТАБ (20,3,2,5,2);
 ТАБ (20,3,2,5,3) - END ⊐
2. Номера столбцов: I,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,
 19,20;
3. Массив ключей: 0,0,0,0,0,0,0,0,0,0,I,I,1,I,I,I,I,I,I;
4. Целевой признак ($T_{л+1}$): 54, 38, 14, 4, I;
5. Первый класс
 (построчно): I,5,7,I8,2,5,I4,4,8,IOI,2,4,3,2,3,I,2,0,I,3,
 2,0,4,II2,4,5,IO,II,I,8I,3,3,0,2,I,3,2,I3,2,
 2,4,I5,IO,5,0,4,7,8,I50,0,I,3,I,I,2,I,2,3,I;
6. Второй класс
 (построчно): 2,2,0,23,I,2,I9,7,4,I90,3,I,2,2,I,I,2,3,I,I,
 I,I,IO,I2,2,4,IO,6,I,90,I,I,3,2,0,2,I,2,2,3;

На печать выдаются по каждой задаче:

1. Исходные данные; 2. Результаты счета.

Задача 1.

Вначале выдаются исходные данные (№ столбцов, ключ, целевой признак, информация первого класса, информация второго класса).

Затем печатаются результаты счета:

а) номера столбцов

I	2	3	4	5	6	7	8	9	10
II	I2	I3	I4	I5	I6	I7	I8	I9	20

б) целевой признак (x_{n+1})

54	38	I4	4	I
----	----	----	---	---

Исходная закодированная таблица

I класс

0	I	I	I	0	I	0	0	I	0	0	0	0	I	0	0	0	0	0	0
I	0	I	I	I	I	0	I	0	0	I	0	0	I	0	I	0	0	I	I
I	I	I	0	I	0	0	I	I	0	0	I	0	0	0	0	I	I	I	0

II класс

I	0	0	I	0	0	I	I	0	I	I	I	I	I	0	0	0	I	0	0
0	0	I	I	0	0	0	I	0	0	I	I	0	I	I	0	I	I	0	0

Информационные веса P_i^v

I	2	3	4	5	6	7	8	9	10
0,166	0,666	0,500	-0,333	0,666	0,666	-0,500	-0,333	0,666	-0,500
II	I2	I3	I4	I5	I6	I7	I8	I9	20
-0,666	-0,666	-0,500	-0,333	-0,500	0,333	-0,166	-0,666	0,666	0,333

Задача 2.

Печатаются исходные данные (№ столбцов, ключ, целевой признак, информация первого класса, информация второго класса). Затем печатаются результаты счета.

а) номера столбцов

I	2	3	4	5	6	7	8	9	10
II	I2	I3	I4	I5	I6	I7	I8	I9	20

б) целевой признак (x_{n+1})

54	38	I4	4	I
----	----	----	---	---

Исходная закодированная таблица.

I КЛ.	0	I	0	I	I	I	0	0	I	0	0	0	0	I	I	0	I	0	0	0
	I	0	0	I	I	I	0	I	0	0	I	0	0	I	0	0	I	0	0	0
	I	I	I	0	I	0	0	I	I	0	0	I	0	0	0	I	0	I	0	I
II КЛ.	I	0	0	I	0	0	I	I	0	I	I	I	I	I	0	0	I	I	0	I
	0	0	I	0	I	0	0	I	0	0	I	I	0	I	0	I	0	I	I	0

Информационные веса ψ_i

I	2	3	4	5	6	7	8	9	10
-28,25	98,25	-62,33	62,33	61,66	I24,0	-61,6	-94,4	98,25	-61,6
II	I2	I3	I4	I5	I6	I7	I8	I9	20
-98,25	I24,0	-61,6	-3,83	94,4	I-62,3	62,33	-I24,0	-66,1	-57,8

Задача 3

Печатаются исходные данные (№№ столбцов, ключ, целевой признак, информация первого и второго классов). Затем печатаются результаты счета.

№№ столбцов.

I	2	3	4	5	6	7	8	9	10
II	I2	I3	I4	I5	I6	I7	I8	I9	20

Целевой признак

54	38	I4	4	I
----	----	----	---	---

Исходная закодированная таблица

I КЛ.	0	I	0	I	I	I	I	0	I	0	I	0	0	I	I	0	I	0	0	0
	I	0	0	I	I	I	0	I	0	0	0	0	0	I	0	0	I	0	0	0
	I	0	I	0	I	0	0	I	I	I	0	I	0	0	0	I	0	I	0	I
II КЛ.	I	0	0	I	0	0	I	I	0	I	0	I	I	I	0	0	I	I	0	I
	0	0	I	0	I	0	0	I	0	0	0	I	0	I	0	I	0	I	I	0

Информационные веса ψ_i^*

I	2	3	4	5	6	7	8	9	10
-22,1	94,0	-85,5	85,5	41,41	I26,9	-52,6	-94,0	80,41	-55,0
II	I2	I3	I4	I5	I6	I7	I8	I9	20
94,0	-I27	-41,4	I3,58	94,0	-85,5	85,5	-I26,9	-71,9	-55,0

ТЕКСТ ПРОГРАММЫ П4

1. BEGIN
2. PROCEDURE TAB (N,K1,K2,K,KЛЮЧ) ; INTEGER N,K1,K2,K, KЛЮЧ ;
3. BEGIN INTEGER ARRAY HOM [1:N] , KЛ [1:N] , И M1, И M2 [1:K] , TK1 [1:K1,1:N] ,
4. TK2 [1:K2,1:N] ; ARRAY BM1,BM2,BM3,И [1:K] , T1 [1:K1,1:N] , T2 [1:K2,1:N] ,
5. BEC [1:N] ; INTEGER I,J,Q,L,S,KM1,KM2,R;
6. REAL BП1, BП2,BП3;
7. PROCEDURE CP; BEGIN
8. FOR J:=1 STEP 1 UNTIL Q DO FOR S:=1 STEP 1 UNTIL K DO
9. IF ABS(BM2[J]-BM1[S])<.10-8 THEN BEGIN И M2[S]:=BM3[J] ;
10. IF KM1=0 THEN И M1 [S] :=BM3 [J] END ;
11. IF KЛЮЧ=1 THEN
12. BEGIN INTEGER I,J,Q;Q:=J:=0; FOR I:=1 STEP 1 UNTIL K1 DO J:=J+ИM2 [I] ;
13. FOR I:=K1+1 STEP 1 UNTIL K DO Q:=Q+ИM2 [I] ; BП2:=J/K1-Q/K2 END ;
14. IF KЛЮЧ=2 THEN
15. BEGIN INTEGER I, J; REAL BП4,BП5, BП6;BП4:=0 FOR I:=1 STEP 1 UNTIL K1 DO BП4:=BП4+1/I;
16. BП5:=0; FOR I:=1 STEP 1 UNTIL K1 DO BEGIN
17. BП6:=0; FOR J:=K1+1 STEP 1 UNTIL K DO

18. $B \Pi 6 := B \Pi 6 + (U \Pi 2 [I] - U \Pi 2 [J]) \times ABS (U [I] - U [J])$;
19. $B \Pi 5 := B \Pi 5 + B \Pi 4 \times B \Pi 6$; $B \Pi 4 := B \Pi 4 - 1/I$ END ; $B \Pi 2 := B \Pi 5 / K2$ END ;
20. IF $K \Pi 04 = 3$ THEN
21. BEGIN INTEGER I, J, Q; REAL B Π 4, B Π 5; B Π 5 := 0;
22. FOR I := 1 STEP 1 UNTIL K-1 DO FOR J := I+1 STEP 1 UNTIL K DO
23. BEGIN B Π 4 := 0; FOR Q := 1 STEP 1 UNTIL J-1 DO B Π 4 := B Π 4 + 1 / (Q \times (K-Q)) ;
24. B Π 5 := B Π 5 + B Π 4 \times (U Π 2 [I] - U Π 2 [J]) \times ABS (U [I] - U [J]) END ; B Π 2 := B Π 5 END ;
25. IF $K \Pi 1 = 0$ THEN B Π 1 := B Π 2;
26. BEGIN INTEGER I ; IF ABS(B Π 2) > ABS(B Π 1) THEN
27. BEGIN B Π 1 := B Π 2; FOR I := 1 STEP 1 UNTIL K DO U Π 1 [I] := U Π 2 [I] END END ;
28. IF $K \Pi 1 = 0$ THEN $K \Pi 1 := 1$ END ;
29. PROCEDURE ΠE (BM, KZ, X); INTEGER X;
30. BEGIN ARRAY BM1 [1:N]; INTEGER I, J;
31. FOR I := 1 STEP 1 UNTIL KZ DO
32. BEGIN FOR J := 1 STEP 1 UNTIL N DO BM1 [J] := BM [I, J];
33. IF X = 0 THEN OUTPUT ('E', BM1, '/') ELSE OUTPUT ('YBD', BM1, '/') END ;
34. OUTPUT ('2/') END ;
35. INPUT(HOM, KJ, U, T1, T2); OUTPUT('Y4B4D', HOM, '2/'); OUTPUT('YBD', KJ, '2/');
36. OUTPUT ('E', U, '2/'); ΠE (T1, K1, 0); ΠE (T2, K2, 0); I := 1;

```

37. M1: FOR J:=1 STEP 1 UNTIL K1 DO BM1 [J] :=T1 [J,I] ;
38. FOR J:=1 STEP 1 UNTIL K2 DO BM1 [K1+J] :=T2 [J,I] ;
39. Q:=0 FOR L:=1 STEP 1 UNTIL K DO BM3 [L] :=BM1 [L] ;
40. FOR L:=1 STEP 1 UNTIL K DO
41. BEGIN B Π3:=BM3 [1] ; FOR J:=2 STEP 1 UNTIL K DO B Π3:=MAX(B Π3,BM3 [J] ) ;
42. IF B Π3> -5107 THEN BEGIN Q:=Q+1;BM2 [Q] :=B Π3 END ;
43. FOR J:=1 STEP 1 UNTIL K DO IF ABS(B Π3-BM3 [J] )<10-8 THEN BM3 [J] :=-108 END ;
44. IF KJ [I] =1 THEN GOTO M3;
45. M9:KM1:=0 FOR L:=1 STEP 1 UNTIL K DO BM3 [L] :=0;L:=1;
46. M2:BM3 [L] :=1;CP; IF L<Q THEN BEGIN L:=-L+1; GOTO M2 END ; GOTO M8;
47. M3:KM1:=0;B Π3:=Q/2+10-8;R:=ENTIER (B Π3);
48. IF R=0 THEN GOTO M9;L:=1;
49. M4: BEGIN INTEGER ARRAY U M3 [1:L] ;
50. FOR J:=1 STEP 1 UNTIL L DO U M3[J] :=J;
51. M5: FOR J:=1 STEP 1 UNTIL Q DO BM3 [J] :=0;
52. FOR J:=1 STEP 1 UNTIL L DO BEGIN S:=U M3 [J] ;BM3 [S] :=1 END ;CP;
53. IF L=1 THEN BEGIN IF U M3 [1] =Q THEN BEGIN GOTO M7 END ELSE BEGIN U M3[1]:=U M3[1]+1;
54. GOTO M5 END END ;J:=1;
55. M6: IF J<L THEN

```

```

56. BEGIN KM2:=1; FOR S:=J+1 STEP 1 UNTIL L DO IF U[M3[S]≠Q-L+S THEN KM2:=KM2*0;
57. IF U[M3[J]<Q-L+J THEN
58. BEGIN IF KM2=1 THEN
59. BEGIN U[M3[J]:=U[M3[J]+1; FOR S:=J+1 STEP 1 UNTIL L DO U[M3[S]:=U[M3[J]+S-J;
60. GOTO M5 END ELSE
61. BEGIN J:=J+1; GOTO M6 END END ELSE GOTO M7 END ELSE
62. BEGIN U[M3[L]:=U[M3[L]+1; GOTO M5 END ;
63. M7: IF L<R THEN BEGIN L:=L+1; GOTO M4 END END ;
64. M8: FOR J:=1 STEP 1 UNTIL K1 DO TK1[J,I]:=U[M1[J];
65. FOR J:=1 STEP 1 UNTIL K2 DO TK2[J,I]:=U[M1 [J+K1]; BEC[I]:=B Π1;
66. IF I<N THEN BEGIN I:=I+1; GOTO M1 END ;
67. OUTPUT('Y4B4D',HOM, '2 /'); OUTPUT('E', U, ' 2/'); Π E(TK1,K1,1) ;
68. Π E(TK2,K2,1); OUTPUT('E',BEC, '4 /') END ;
69. TAB (20,3,2,5,1); TAB (20,3,2,5,2); TAB (20,3,2,5,3) END

```

ПРОГРАММА П5

"Оптимальное бинарное кодирование признаков, подсчет строчечных нагрузок и минимизация" составлена на языке "Альфа" для ЭВМ М-222.

А.Д.Соколов

Назначение. Программа П5 предназначена для оптимального бинарного кодирования признаков, заданных в количественном, порядковом и логическом видах. После кодирования производится подсчет строчечных нагрузок и минимизация признакового пространства. Минимизация производится с учетом "идеальных" объектов. Программа может быть использована для двухклассовых постановок задач.

Порядок постановки перфокарт.

1. Система "Альфа" (№ I25 п/к)
2. M - число количественных признаков, $k\Sigma$
3. MI - общее количество признаков, $k\Sigma$
4. J - число объектов I класса, $k\Sigma$
5. J - число объектов II класса, $k\Sigma$
6. N - номера признаков, $k\Sigma$
7. Массив A (табл.), $k\Sigma$ - } (числа) каждый признак и каждый мас-
8. Массив B (табл.), $k\Sigma$. } сив разделяется $k\Sigma$ и 2 пустыми п/к .

Данная программа позволяет обрабатывать таблицы с $(KI+K2)=200$ и $n2=250$. На печать выдается:

1. N - номер признака (N раз)
2. B - граничные числа ($N+1$ раз)
3. E - вес признака (M раз)
4. Порядковые номера столбцов после упорядочивания по P_j^v
5. Пошаговое минимальное значение строчечной нагрузки $J(P_j^v)$ для I кл.
6. Пошаговое максимальное значение строчечной нагрузки $J(P_j^v)$ для II класса.
7. Пошаговая величина "растяжки" классов ρ^1 . Все эти величины выдаются N раз.
8. Строчечные нагрузки объектов I и II классов на полном пространстве признаков ($J+J$ раз).
9. $I/2$ значения строчечной нагрузки "идеального" объекта I класса ($RI/2$) на полном пространстве признаков (I раз).

10. Ширина зоны неопределенности на полном пространстве признаков (I раз).
11. I/2 значения строчечной нагрузки "идеального" объекта I класса на минимизированном пространстве признаков (I раз).
12. Ширина зоны неопределенности на минимизированном пространстве признаков (I раз).
13. Номер последнего признака, вошедшего в минимизированный набор признаков (I раз).
14. Строчечные нагрузки объектов I и II классов в минимизированном пространстве признаков (J + J раз).

Контрольный пример.

Имеется таблица, состоящая из 2-х классов:

I кл. 3 строки и 4 столбца

II кл. 2 строки и 4 столбца

№	I	2	3	4	
1	15	1	1	10	I кл.
2	20	1	0	8	
3	15	3	0	8	
4	18	1	0	2	II кл.
5	10	2	0	6	

Закодированная информация имеет вид:

κ	КОП	A1	A2	A3	
+	01	200	-	-	перфокарта M
+	01	200	-	- KΣ	

κ	КОП	A1	A2	A3	
+	01	400	-	-	перфокарта MI
+	01	400	-	- KΣ	

κ	КОП	A1	A2	A3	
+	01	300	-	-	перфокарта J
+	01	300	-	- KΣ	

κ	КОП	A1	A2	A3	
+	01	200	-	-	перфокарта J
+	01	200	-	-	

κ	КОП	A1	A2	A3	
+	01	100	-	-	перфокарта N
+	01	400	-	-	
+	01	200	-	-	
+	01	300	-	-	

Таблица

κ	КОП	A1	A2	A3		
+	02	150	-	-	+KΣ	признак № 1
+	02	200	-	-		
+	02	150	-	-		
+	02	180	-	-	+KΣ	
+	02	100	-	-		
+	02	100	-	-	+KΣ	признак № 4
+	01	800	-	-		
+	01	800	-	-		
+	01	200	-	-	+KΣ	
+	01	600	-	-		
+	01	100	-	-	+KΣ	признак № 2
+	01	100	-	-		
+	01	300	-	-		
+	01	100	-	-	+KΣ	
+	01	200	-	-		
+	01	100	-	-	+KΣ	признак № 3
+	00	000	-	-		
+	00	000	-	-		
+	00	000	-	-	+KΣ	

Кодирование (а следовательно, и перфорация) признаков производится в определенной последовательности: сначала кодируются количественные признаки, затем - дискретные.

Причем, перфорируется признак в I-м классе, затем находится его сумма (KΣ), а после этого перфорируется этот же признак во

П-м классе и находится его сумма и т.д. Результаты счета представлены в следующем виде:

I. Кодирование и подсчет информационных весов признаков

Кодирование методом "скользящей границы"

№ признака	Граничное число В	Информационный вес Е
$n = 1$	$B = 12,5$	$E = - 0,50$
$n = 4$	$B = 6,0$	$E = - 1,00$

Кодирование методом "перебора"

№ признака	Граничное число В	Информационный вес Е
$n = 2$	$B = 1,0$	$E = 0,33$
$n = 3$	$B = 1,0; 3,0$	$E = + 0,50$

№№ признаков, упорядоченные по убыванию значения P_j^v
 4 3 1 2

II. Подсчет строчечных нагрузок и минимизация признакового пространства

	$\min J(P_j^v)^I$	$\max J(P_j^v)^{II}$	β^L - величина "растяжки" классов		
I шаг	1,0	0,0	1,0		
II шаг	1,5	0,5	1,0		
III шаг	2,0	1,0	1,0		
IV шаг	2,0	1,0	1,0		
$J(P_j^v)$ - на полном пространстве признаков - I класс					
	$J(P_1^v)$	$J(P_2^v)$	$J(P_3^v)$	$J(P_4^v)$	$J(P_5^v)$
	2,3	2,0	2,0	1,0	0

$R_1/2$ - на полном пространстве признаков = + 1,166

Ширина зоны неопределенности на полном пространстве признаков
 = + 0,333

$R_1/2$ на минимизированном пространстве признаков + 0,500

Ширина зоны неопределенности на минимизированном наборе признаков. ± 0,000

№ последнего признака, вошедшего в минимизированный набор

признаков $N = 4$

Строчечные нагрузки объектов I и II классов в минимизированном пространстве признаков

$J(P_1^v)$	$J(P_2^v)$	$J(P_3^v)$	$J(P_4^v)$	$J(P_5^v)$
1,00	1,00	1,00	0,00	0,00

ТЕКСТ ПРОГРАММЫ П5

начало целый p3, p4; вещ a7; вещ a5; вещ a1, в1, в, a2, a3, a4; вещ B2; целый m1;
целый i, j, k, l, m, n, p1, p2, J, J; вещ a, в, с, d, e; логич *альфа *;
массив A, A1, B, B1[1:100], U[1:250]; массив P1[1:250]; массив A2[1:250];
логич массив L [1:250, 1:200]; процедура БОРТ (X, K); начало целый k, l; вещ x;
l:=k; M:l := l - 1; для k:=1, ..., l цикл начало если X[k] > X[k+1] то начало x:=X[k+1];
X[k+1] :=X[k]; X[k]:=x конец конец; если l>1 то на M; конец; ввод (m1); n :=0;
ввод (I, J, m); СПО176(1072, P1[1], P1[m], 0, 0); Щ:n :=n + 1; СПО176 (1072, A[1], A[I], 0, 0);
СПО176 (1072, B[1], B[J], 0, 0); p1:=p2:=0; c:=0; a:=0; если n>m то на Щ1; для i :=1, ..., I
цикл A1[i] :=A[i]; для j:=1, ..., J цикл B1[j]:=B[j]; БОРТ (A, I); БОРТ (B, J); Я:k:=k+1;
если abs (A1[1] - B[1]) <₁₀-5 то начало p3:=0; p4:=0; для i:=1, ..., I-p1 цикл если
abs (A[i]-A[i]) <₁₀-5 то p3:=p3+1; для j :=1, ..., J -p2 цикл если abs(B[1] - B[j]) <₁₀-5
то p4:=p4+1; p1:=p1+p3; p2:=p2+p4; a:=A[1]; для i :=1, ..., I-p1 цикл A[i] :=A[i]+p3;
для j := 1, ..., J-p2 цикл B[j]:=B[j]+p4; a1:= если A[1] < B[1] то A[1] иначе B[1];
на Mu конец; если A[1] < B[1] то начало p3:=0; для i := 1, ..., I-p1 цикл если
abs (A[i] - A[i]) <₁₀-5 то p3:=p3+1; p1:=p1+p3; a:=A[1]; для i:= 1, ..., I-p1 цикл
A[i] :=A[i]+p3; a1 := если A[1] < B[1] то A[1] иначе B[1]; конец иначе начало p4:=0;
для j := 1, ..., J -p2 цикл если abs(B[1]-B[j]) <₁₀-5 то p4:=p4+1; p2:=p2+p4; a:=B[1];
для j := 1, ..., J -p2 цикл B[j] :=B[j]+p4; a1:= если A[1] < B[1] то A[1] иначе B[1];
конец; Mu := U [n] :=p1/I-p2/J; a:= U [n]; если abs(d) > c то начало c:= abs(d);
e:=d; b :=(a+a1)/2 конец; если p1 = 1 и p2=J то на B; на Я; B: вывод (истина, P1[n], в, e,
ложь); U [n] :=e; для i := 1, ..., I цикл начало A1[i] := если A1[i]-B <₁₀-4 то 1 иначе 0;
L[n, i] := если abs(A1[i]- 1) <₁₀-4 то истина иначе ложь; конец; для j := 1, ..., J цикл
начало B 1[j] := если B1[j] - в <₁₀-4 то 1 иначе 0; L[n, I+j] := если abs (B1 [j] - 1) <₁₀-4
то истина иначе ложь; конец; если n<m то на Щ; если n=m то на Щ2; Щ1 : A2 [] :=A [];

OPT (A,I); U[n]:=0; A[1]:=V1[] :=0; Я1 : р3:=0; р4:=0; для i := 1 ,..., I-р1 цикл если
 abs(A[1] - A [i]) < 10^{-5} то р3:=р3+ 1; для j:= 1, ..., J -р2 цикл если abs (A[1] -B[j]) < 10^{-5}
то р4:=р4+1 ; а:=A [i] ; M_и 1: d:=р3/I-р4/J ; если d>0 то начало U[n]:= U[n]+ d ; для j:=
 1 ,..., J цикл если abs (B[j] -a) < 10^{-5} то B1[j] :=1 ; для i:= 1, ..., I цикл если abs (A2[i]
 -a) < 10^{-5} то A1[i]:= 1; вывод (а, истина) конец ; р1:=р1+ р3; для i:=1 ,..., I-р1 цикл A[i]
 :=A[i]+р3 ; если р1=I то на Я1 ; на Я1 ; Я1 : вывод (P1[n] , истина) ; вывод (U[n] ,
 истина); для i:= 1, ..., I цикл Л[n,j]:=если abs(A1 [i] - 1) < 10^{-4} то истина иначе ложь; для
 j :=1 ,..., J цикл Л[n, I+ j]:= если abs (B1 [j] - 1) < 10^{-4} то истина иначе ложь; если
 n<m то на Щ; Щ2: i:=m ; Ю: i:=i -1 ; для j:= 1, ..., i цикл начало еслиabs(U[j])
 abs (U[j + 1]) то начало l :=P1 [j + 1]; а:= U [j+1]; для k:= 1, ..., I+ J цикл начало *альфа*
 := Л [j+1, k] ; Л [j+1 , k] :=Л [j,k] ; Л [j,k] :=*альфа* конец; P1 [j+1]:=P1 [j] ; P1 [j]:= 1;
 U [j+1]:=U [j]; U [j] := а конец конец ; если i>1 то на Ю; СП0176 (1153, P1 [1] , P1 [m] , 0, 0); вывод
 (истина); а5:=0; а2:= 10^6 ; A [] :=B [] :=0; i :=0; Юi : i := i+1 ; для j:=1 ,..., I цикл начало
еслиU [i]>0 and Л [i, j] or U [i]<0 and notЛ [i, j] то A [j]:=A [j] +abs (U [i]) конец; для j:=
 1 ,..., J цикл начало еслиU [i]>0 and Л [i, I+j] or U [i]<0 and notЛ [i, I+j] то B [j] :=B [j]
 +abs (U [i]) конец ; а5:=а5+abs (U [i]); v1:=а5/2; а:= 10^6 ; в:=0; для j:=1 ,..., I цикл если
 A [j] < а то а:=A [j] ; для j:=1 ,..., J цикл если B [j] > в то в:=B [j] ; с:=а-в; если а > в
то начало а7:=в+(а-в)/2; в2:=abs (в1-а7) конец; если а < в то начало если в1 > в or abs (в1-в) < 10^{-5}
то в2:=abs (в1-а); если в1 < а or abs (в1-а) < 10^{-5} то в2:=abs (в1-в); если в1 < в and в1 > а то
 в2:=abs (в-а) конец; если abs (а-в) < 10^{-5} то в2:=abs (в1-а); если в2 < а2 то начало а4:=P1 [i] ;
 а2:=в2; а3:=в1; A1 [] :=A [] ; B1 [] :=B [] конец; вывод (а, в, с, истина); если i =m то начало
 СП0176(1153, A [1] , A [I] , 0, 0); вывод (ложь); СП0176(1153), B [1] , B [J] , 0, 0); вывод (ложь);
 вывод (в1, в2, истина) конец ; если i < m то на Ю1 ; вывод (а3, а2, а4, ложь); СП0176(1153, A [1] [1],
 A1 [I] , 0, 0); вывод (ложь); СП0176(1153, B1 [1] , B1 [J] , 0, 0); вывод (ложь); конец *

Таблица распределения памяти

	Начало	Конец	Длина
Программа	0020	2115	2076
Числовые константы	2116	2141	0024
Формируемые константы	2142	2225	0064
Скаляры	2226	2262	0035
Рабочие ячейки	2263	2273	0011
Массивы			
1	2274	2665	0372
2	2666	3031	0144
3	3032	3175	0144
4	3176	3341	0144
5	3342	3505	0144
6	3506	4077	0372
7	4100	4471	0372
IO	4472	6621	2130
Рабочее поле IC	6622	7500	0656
Адреса программных остановов	2115		

ЧАСТЬ IV.

Программы для статистической обработки геологической ин- формации.

Программа П1.

"Главные компоненты"

Составлена на языке Алгол-БЭСМ для ЭВМ БЭСМ-6.

Е.И.Бурова, В.Д.Карбышев.

Назначение.

В геологии при изучении объектов, охарактеризованных многими признаками, используются различные методы многомерной статистики среди которых определенное место занимает метод главных компонент (МГК).

МГК может использоваться для решения задачи выявления факторов (процессов), влияющих на формирование геологического объекта на основании наблюдений над современным обликом геологических объектов и исследований фактических значений их параметров. Одна из трудностей решения такой задачи заключается в том, что доступные для наблюдения значения тех или иных признаков, представляя собой конечный результат совместной деятельности целой совокупности факторов (процессов), которые могут оказывать совместное и различное по интенсивности влияние на значения признаков. Решение задачи возможно, если факторы характеризуются своим, присущим только им, набором значений признаков, либо различными направленностью и интенсивностью воздействия на них. В качестве диагностических характеристик (для выявления факторов) выступают взаимосвязанные изменения признаков изучаемой совокупности [5] .

В основу МГК положено исследование внутренней структуры корреляционной или ковариационной матрицы, в результате которого находится такое линейное преобразование заданных m случайных величин (признаков x_1, \dots, x_m) в новый набор P случайных величин (главные компоненты или факторы), которые статистически независимы между собой и расположены в порядке уменьшения собственных значений соответствующей матрицы [6] . В геометрической интерпретации главные компоненты (факторы) представляют оси координат m -мерного эллипсоида, образованного значениями соответствующих признаков и являются своеобразной характеристикой изменчивости

в пространстве значений m - признаков в этих направлениях (вдоль осей координат). Факторные нагрузки (собственные вектора) представляют собой долю участия каждого признака в этих направлениях или еще, как коэффициент корреляции между соответствующими главной компонентой и признаком.

Краткое описание алгоритма.

§ I. Постановка задачи.

Задава таблица объектов, строками которой являются объекты, а столбцами - признаки.

На пересечении строки i и столбца j в таблице находится значение, которое принимает j -ый признак на i -ом объекте. При мером может служить такая таблица.

Таблица I.

№ объекта	признак	признак	признак	признак
	1	2	3	4
1	1	1	2	3
2	2	2	3	4
3	3	1	4	1
4	4	2	2	2
5	1	1	3	3
6	2	2	4	4
7	3	1	3	1
8	4	2	3	2

Обозначим через КОБ - количество объектов и через КП - количество признаков в таблице. Требуется распечатать таблицу объектов. Требуется посчитать и выдать на печать выборочное среднее каждого признака по формуле: $[I]$.

$$\frac{1}{\text{КОБ}} \sum_{i=1}^{\text{КОБ}} x_{ij} = \bar{x}_j \quad j = 1, 2, \dots, \text{КП}$$

Здесь x_{ij} есть значение признака j на i -ом объекте. Требуется посчитать и выдать на печать выборочную дисперсию каждого признака по формуле:

$$\sqrt{\frac{1}{\text{КОБ} - 1} \sum_{i=1}^{\text{КОБ}} (x_{ij} - \bar{x}_j)^2} = \sigma_j ; \quad j = 1, 2, \dots, \text{КП}$$

Требуется посчитать и выдать на печать выборочную матрицу ковариаций, каждый элемент которой определяется формулой:

$$\frac{1}{\text{КОБ}-1} \sum_{i=1}^{\text{КОБ}} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = \sigma_{jk}; \quad j, k = 1, 2, \dots, \text{КП}.$$

Требуется посчитать и выдать на печать выборочную матрицу корреляций между признаками, элементы которой описываются формулой:

$$\frac{\sigma_{jk}}{\sigma_j \cdot \sigma_k} = \rho_{jk} \quad j, k = 1, 2, \dots, \text{КП}$$

Обозначим через σ сумму

$$\sigma = \sum_{j=1}^{\text{КП}} \sigma_{jj}$$

А: требуется найти и распечатать s собственных значений $\lambda_1 > \lambda_2 > \dots > \lambda_s$ матрицы ковариаций таких, что $\sum_{k=1}^s \lambda_k \geq \sigma \cdot \text{Порог}$ и $\sum_{k=1}^{s-1} \lambda_k < \sigma \cdot \text{Порог}$, где параметр $0 \leq \text{Порог} \leq 1$.

В § 18,6 такая система собственных значений определяется как главные компоненты. [1].

В: требуется найти и распечатать соответствующие этим собственным значениям собственные вектора $C_{11}, C_{22}, \dots, C_{ss}$ матрицы ковариаций с координатами в исходном признаковом пространстве равными:

$$\begin{aligned} & (C_{11}, C_{12}, \dots, C_{1\text{КП}}) \\ & (C_{21}, C_{22}, \dots, C_{2\text{КП}}) \\ & \dots \dots \dots \\ & (C_{s1}, C_{s2}, \dots, C_{s\text{КП}}) \end{aligned}$$

и такие, что для всякого $k = 1, 2, \dots, s$ выполняется равенство:

$$\sum_{j=1}^{\text{КП}} C_{kj}^2 = 1$$

Система собственных векторов представляет ортонормированный базис.

Требуется найти и распечатать координаты объектов в новом признаковом пространстве, задаваемом найденным базисом, используя их координаты в исходном признаковом пространстве, задаваемые таблицей объектов.

Координаты в новом пространстве находятся по формуле:

$$x'_{ik} = \sum_{j=1}^{\text{КП}} C_{kj} \cdot x_{ij}$$

где: x'_{ik} - новая координата i -го объекта по k -му собственному вектору; C_{kj} - j -ая координата k -го собственного вектора в исходном признаковом пространстве; x_{ij} определяется выше.

Если вместо таблицы объектов требуется обработать уже из-

вестную задаваемую пользователем матрицу ковариаций по пунктам А и В, описанным выше, то программа должна это предусматривать.

Замечание: в реализованной программе в этом режиме можно обрабатывать любую симметричную не вырожденную матрицу с положительными диагональными элементами.

Если диагональные элементы исходной матрицы ковариаций окажутся нулевыми, то программа должна распечатывать контекст:

"ДИАГОНАЛЬНЫЕ ЭЛЕМЕНТЫ НУЛЕВЫЕ. ЗАДАЧА БЕССМЫСЛЕННА"

и перейти к решению следующего варианта, если таковой имеется.

Если исходная матрица ковариаций окажется несимметричной, то программа должна распечатать такой контекст:

"МАТРИЦА НЕСИММЕТРИЧНА"

и перейти к решению следующего варианта, если таковой имеется.

Требуется обеспечить многовариантную работу программы.

§ 2. Описание метода поиска собственных векторов.

В программе реализован итерационный метод поиска собственных значений и собственных векторов симметричной матрицы, описанный в [2] , гл.8, § 7. Этот метод позволяет искать собственные значения матрицы в порядке их убывания и прервать этот поиск по достижению заданного порога от ϵ .

Точность, с которой в программе находится значение отдельного собственного значения, регулируется параметром ТОЧНОСТЬ.

Расчет отдельного собственного значения прерывается, если выполняется неравенство:

$$|\lambda_{n+1} - \lambda_n| < \text{ТОЧНОСТЬ},$$

где λ_n - собственное значение, вычисленное на итерации n , а λ_{n+1} - тоже собственное значение, вычисленное на итерации $n+1$.

ИНСТРУКЦИЯ К ПОЛЬЗОВАНИЮ.

Программа предназначена для расчетов на ЭВМ "БЭСМ-6" с использованием транслирующей системы АЛГОЛ-БЭСМ.

Комплектация колоды.

На карте 1 должен быть пробит шифр задачи в стандартном виде. Например, шифр 205015 пробивается в виде

ШИФР 205015

На карте 2 должно быть пробито требование на время в стан-

дартном виде. Например, требование в 2 часа 32 минуты 00 сек. пробивается в виде:

ВРЕМЯ 023200

Между картами 82 и 83 должны быть вставлены карты с обращениями к процедуре обработки. Обращение имеет вид

ОСНОВНАЯ (КОБ, КП, ТОЧНОСТЬ, ПОРОГ)

КОБ = I, если производится обработка заданной пользователем симметричной матрицы;

КОБ равно количеству объектов в таблице объектов и всегда должно быть строго больше I, если обрабатывается таблица объектов;

КП - равно количеству признаков в таблице объектов (матрице ковариаций);

ТОЧНОСТЬ равна абсолютной погрешности, с которой должны считаться собственные значения;

ПОРОГ равен величине порога от 6 .

Если предстоит расчет нескольких вариантов, должно быть несколько обращений к процедуре обработки, разделенных точкой с запятой. После последнего обращения к процедуре обработки точка с запятой не ставится, пробивается служебное слово END по правилам принятым в системе АЛГОЛ-БЭСМ и символ конца строки с кодом I74.

На стр.170 в строках 7) и 8) приведен пример 4-кратного обращения к процедуре. Между картами 85 и 86 должны быть вставлены карты с числовыми массивами. Один массив соответствует одному обращению к процедуре обработки и количество массивов должно быть равно числу обращений к процедуре обработки. Внутри массивов числа отделяются друг от друга запятыми. После каждого массива ставится точка с запятой.

Отдельный массив представляет либо запись таблицы объектов, либо запись симметричной матрицы. Как таблицы, так и матрицы выписываются последовательно строка за строкой, начиная с I-ой строки. Числа в строках выписываются в порядке слева направо. Количество чисел в массиве должно в точности совпадать с количеством элементов в таблице (матрице), которое может быть вычислено по виду обращения к процедуре обработки. А именно: если обрабатывается симметричная матрица, то число элементов в массиве должно быть равно $KП \times KП$; если обрабатывается только таблица объектов, то число элементов в массиве должно быть $КОБ \times KП$.

Замечание. Несоблюдение последнего правила в программе не

контролируется, счет будет выполнен и в результате могут быть получены неверные ответы.

Каждое обращение к процедуре обработки работает на свой массив данных. 1-е обращение на 1-ый по счету массив, 2-е обращение на 2-ой по счету массив и т.д.

На стр. 170 в строках с 1) по 6) закодированы следующие данные, соответствующие четырем обращениям к процедуре обработки, написанных в строках 7 и 8.

Перфорацию следует проводить на устройстве УЩ.

Таблица 2

0	I	2
I	0	3
2	3	0

Таблица 3.

I	2	3	4
2	I	3	4
3	3	I	4
4	4	2	I

Таблица 5

I	I	2	3
2	2	3	4
3	I	4	I
4	2	2	2
I	I	3	3
2	2	4	4
3	I	3	I
4	2	3	2

Таблица 4

I	0,438	-0,137	0,205	-0,178
0,438	I	0,03I	0,180	-0,304
-0,137	0,03I	I	0,16I	0,372
0,205	0,180	0,16I	I	-0,013
-0,178	-0,304	0,372	-0,013	I

Эти данные закодированы так:

- 1) 0, I, 2, I, 0, 3, 2, 3, 0; I, 2, 3, 4, 2, I, 3, 4, 3, I, 4, 4, 4, 2, I;
- 2) I, 0.438, -0.137, 0.205, -0.178, 0.438, I, 0.03I, 0.180, -0.304,
- 3) -0.137, 0.03I, I, 0.16I, 0.372, 0.205, 0.180, 0.16I, I, -0.013,
- 4) -0.178, -0.304, 0.372, -0.013, I;
- 5) I, I, 2, 3, 2, 2, 3, 4, 3, I, 4, I, 4, 2, 2, 2, I, I, 3, 3, 2, 2, 4, 4, 3, I, 3, I,
- 6) 4, 2, 3, 2;

Информационные карты.

- 7) ОСНОВНАЯ (I, 3, I₀-3, 0.95); ОСНОВНАЯ (I, 4, I₀-3, 0.95);
- 8) ОСНОВНАЯ (I, 5, I₀-5, 0.99); ОСНОВНАЯ (8, 4, I₀-5, 0.9) - END

Пояснения к результатам

решения контрольного примера (4 теста).

Выдача с распечаткой программы на АЛГОЛЕ и результатами решения тестовых вариантов.

1-ый тест представляет обработку массива чисел таблицы 2,

2-ой тест - таблицы 3,

3-ий тест - таблицы 4, которая заимствована из [3], стр. 61, таблица 4.3.

Расчет был проведен с абсолютной погрешностью 10^{-5} , т.е. более точно, чем это сделано в [3]. Поэтому собственные значения в тесте совпадают с собственными значениями в [3], если первые округлить до третьего знака.

Для совпадения координат собственных векторов (компоненты) приведенных в таблице 4.4 в [3] на стр. 61, с координатами, полученными в тесте, необходимо умножить координаты каждого собственного вектора, приведенные в тесте, на квадратный корень из соответствующего ему собственного значения.

Это связано с тем, что в программе ищется ортонормированный базис из собственных векторов, в то время как компоненты в [3] таковыми не являются.

4-ый тест представляет результат обработки таблицы 5.

Полученная в этом случае матрица ковариаций оказалась вырожденной (в данном примере для третьего признака). В этом случае [1], как известно, одно из собственных значений должно быть равно 0. Такие собственные значения настоящей программой найдены быть не могут. Поэтому следует избегать обработки настоящей программой вырожденных симметричных матриц.

Выдача на печать:

1. Входная матрица - таблица 2.

2. Диагональные элементы нулевые. Задача бессмысленна.

3. Входная матрица - таблица 3.

4. Матрица не симметрична.

5. Входная матрица - таблица 4.

6. Главные компоненты λ_s .

1) 1,757 ; 2) 1,331 ; 3) 0,781 ; 4) 0,709 ; 5) 0,422.

5. Координаты собственных векторов C_s .

C 1 : 0,556; 0,565; -0,268; 0,237; - 0,493

C 2 : 0,184; 0,245; 0,663; 0,555; 0,398

С 3 : 0,230; 0,432; 0,310; - 0,788; 0,207.

С 4 : 0,634; - 0,318; -0,408; 0,005; 0,575.

С 5 : -0,448; 0,577; -0,476; 0,122; 0,473.

6. Таблица объектов - таблица 5.

7. Выборочное среднее и дисперсия для каждого признака.

8. Выборочная матрица ковариаций

1,428	0,286	0	-0,857
0,285	0,286	0	0,286
0	0	0,571	0
-0,851	0,286	0	0,143

Матрица ковариаций вырожденная (для 3-го признака) см. пояснение к результатам решения (стр. 171).

9. Главные компоненты, координаты собственных векторов, таблица объектов в новой системе координат в данном случае не имеют смысла.

Л И Т Е Р А Т У Р А

1. Уэлкс С. "Математическая статистика". Изд-во "Наука", 1967.
2. Березин И.С., Жидков Н.П. "Методы вычисления". Изд-во "Физматгиз", 1962, т. № 1.
3. Д. Лоули, А. Максвелл "Факторный анализ как статистический метод", изд-во "Мир", 1967, перевод с английского.
4. Курочкин В.М., Подшивалов Д.Б. и др. "Система БЭСМ-АЛГОЛ. Методическое руководство по программированию". ВЦ МГУ, М., 1969.
5. М.Д. Беловин и др. "Факторный анализ в нефтяной геологии" ВИЭСМ, Москва, 1971.
6. У. Крамбейн и др. "Модели геологических процессов". Изд-во "Мир", М., 1973.

```

1. BEGIN
2. PROCEDURE ОСНОВНАЯ( КОБ, КП, ТОЧНОСТЬ, ПОРОГ);
3. INTEGER КОБ,КП; REAL ТОЧНОСТЬ, ПОРОГ;
4. BEGIN INTEGER I,J,K, КЛЮЧ ,S;
5. REAL ВП1, ДИС, ВЕ2, СЧ1, СЧ2;
6. ARRAY ВМ2, ВМ5, ВМ3 [1:КП],СВ, ВМ4 [1:КП, 1:КП ], ВМ1 [1:КОБ, 1:КП];
7. IF КОБ =1 THEN BEGIN INPUT(ВМ4);
8. OUTPUT ( ' Т ' , ' ВХОДНАЯ:МАТРИЦА ' , ' 1/' );
9. FOR I:=1 STEP 1 UNTIL КП DO BEGIN FOR J:=1 STEP 1 UNTIL К П DO
10. ВМ5 [J] :=ВМ4 [ I,J ]; OUTPUT ( ' Е ' , ВМ5, '/' ) END ; OUTPUT ( ' 2/' ) ;
11. FOR I:=1 STEP 1 UNTIL К П DO FOR J:=1 STEP 1 UNTIL К П DO BEGIN
12. IF ВМ4 [ I,J ] ≠ ВМ4 [ J,I ] THEN GOTO М3 END ; GOTO М1 ;
13. М3:OUTPUT ( ' Т ' , ' МАТРИЦА:ВНЕ:ВСИММЕТРИЧНА ' , ' 2/' ) ;
14. GOTO М4 END:
15. INPUT(ВМ1);OUTPUT('Т', ' ТАБЛИЦА:ВОБЪЕКТОВ ' , ' 1/' ) ;
16. FOR I:=1 STEP 1 UNTIL КОБ DO BEGIN FOR J:=1 STEP 1 UNTIL КП DO

```


17. $BM5 [J] := BM1 [I, J];$ OUTPUT ('E', BM5, '/') END ; OUTPUT ('2 /');
18. FOR I:=1 STEP 1 UNTIL КП DO BEGIN ВП 1:=0;
19. FOR J:=1 STEP 1 UNTIL КОБ DO ВП 1:=ВП 1+BM1 [J, I]; ВП 1:=BM2[I]:=ВП 1 / КОБ ;
20. OUTPUT ('T', ' ВБОРОЧНЕ: ВСРЕДНЕЕ: ВПРИЗНАКА', '22В - DDD2В', I,
21. 'T', ' РАВНО', '3В', 'E', ВП 1, '/'); ВП 1:=0;
22. FOR J:=1 STEP 1 UNTIL КОБ DO В П1:=В П1+(BM1 [J, I] - BM2[I])²;
23. В П1:=BM3[I]:=SQRT (В П1 / (КОБ -1));
24. OUTPUT ('T', ' ВБОРОЧНАЯ: ВДИСПЕРСИЯ: ВПРИЗНАКА', '22В - DDD2В', I,
25. 'T', ' РАВНА', '3В', 'E', ВП 1, '1 /') END ;
26. FOR I:=1 STEP 1 UNTIL КП DO FOR J:=1 STEP 1 UNTIL КП DO BEGIN
27. ВП 1:=0; FOR K:=1 STEP 1 UNTIL КОБ DO
28. ВП 1:=ВП 1+(BM1 [K, I] -BM2 [I])*(BM1 [K, J] -BM2 [J]);
29. BM4 [I, J] :=ВП 1 / (КОБ -1) END ;
30. OUTPUT ('T', ' ВБОРОЧНАЯ: МАТРИЦА: КОВАРИАЦИИ', '1 /');
31. FOR I:=1 STEP 1 UNTIL КП DO BEGIN FOR J:=1 STEP 1 UNTIL КП DO
32. BM5 [J] :=BM4 [I, J] ; OUTPUT('E', BM5, '/') END ; OUTPUT('2 /');
33. OUTPUT ('T', ' ВБОРОЧНАЯ: МАТРИЦА: КОРРЕЛЯЦИИ', '1 /');
34. FOR I:=1 STEP 1 UNTIL КП DO BEGIN FOR J:=1 STEP 1 UNTIL КП DO

```

35. BEGIN IF BM3[I]<10-10 THEN BEGIN IF BM3[J]<10-10 THEN BM3[J]:=1
36. ELSE BM5[J]:=0; GOTO M2 END ;
37. IF BM3[J]<10-10 THEN BEGIN IF BM3[I]<10-10 THEN BM5[J] :=1
38. ELSE BM5[J] :=0; GOTO M2 END ;
39. BM5[J] :=BM4 [I,J] / (BM3[I]*BM3[J]);M2: END ;
40. OUTPUT ( 'E',BM5, '/' ) END ;OUTPUT('2 /');
41. M1: ДИС :=0; FOR I:=1 STEP 1 UNTIL KΠ DO ДИС :=ДИС +BM4 [ I,I ] ;
42. IF ДИС <10 -10 THEN BEGIN
43. OUTPUT('T', 'ДИАГОНАЛЬНЫЕ ЭЛЕМЕНТЫ: ВУЗЕБЕ', '/', 'T',
44. 'ЗАДАЧА: ВЕРСОВСКОЕ', '2 /' ) ; GOTO M4 END ; I:=1;
45. M5: BΠ2:=0; FOR J:=1 STEP 1 UNTIL KΠ DO BEGIN BΠ1:=0;
46. FOR K:=1 STEP 1 UNTIL KΠ DO BΠ1:=BΠ1+BM4[K,J];BM3[J] :=BΠ1;BΠ2:=BΠ2+BΠ1 END ;
47. FOR J:=1 STEP 1 UNTIL KΠ DO BM3[J] :=BM3[J]/BΠ2; KΠOЧ:=1;
48. M6: FOR J:=1 STEP 1 UNTIL KΠ DO BEGIN BΠ1:=0;
49. FOR K:=1 STEP 1 UNTIL KΠ DO BΠ1:=BΠ1+BM4 [J,K]*BM3[K];BM5[J] :=BΠ1 END ;
50. BΠ1:=BΠ2:=0; FOR J:=1 STEP 1 UNTIL KΠ DO BEGIN
51. BΠ1:=BΠ1+BM3[J]*BM5[J];BΠ2:=BΠ2+BM3[J]↑2 END ;
52. CУ2:=BΠ1 /BΠ2; IF KΠOЧ≠0 THEN GOTO M7;
53. IF ABS(CУ2-CУ1)<ТОЧНОСТЬ THEN GOTO M8;

```

54. $M7:KJLH:=0;BП1:=0; \text{FOR } J:=1 \text{ STEP } 1 \text{ UNTIL } KП \text{ DO } BП1:=BП1+BM5[J]^2;$
 55. $BП1:=\text{SQRT}(BП1); \text{FOR } J:=1 \text{ STEP } 1 \text{ UNTIL } KП \text{ DO } BM3[J]:=BM5[J] / BП1; CЧ 1:=CЧ2; \text{GOTO } M6;$
 56. $M8:BM2[I]:=CЧ 2;BП1:=0 \text{ FOR } J:=1 \text{ STEP } 1 \text{ UNTIL } KП \text{ DO } BП1:=BП1+BM5[J]^2;$
 57. $BП1:=\text{SQRT}(BП1); \text{FOR } J:=1 \text{ STEP } 1 \text{ UNTIL } KП \text{ DO } CB[I,J]:=BM5[J] / BП1;$
 58. $BП1:=0; \text{FOR } J:=1 \text{ STEP } 1 \text{ UNTIL } I \text{ DO } BП1:=BП1+BM2[J];$
 59. IF $BП1 / \text{ДИС} > \text{ПОРОГ}$ THEN GOTO Н1ХQД;
 60. FOR $J:=1 \text{ STEP } 1 \text{ UNTIL } KП \text{ DO FOR } K:=1 \text{ STEP } 1 \text{ UNTIL } KП \text{ DO}$
 61. $BM4[J,K]:=BM4[J,K] - BM2[I] * CB[I,J] * CB[I,K]; I:=I+1; \text{GOTO } M5;$
 62. Н1ХQД : BEGIN ARRAY $BM6[1:I]; \text{FOR } J:=1 \text{ STEP } 1 \text{ UNTIL } I \text{ DO } BM6[J]:=BM2[J];$
 63. OUTPUT ('T'. 'ГЛАВНЫЕ:РКОМПОНЕНТЫ', '1/');
 64. OUTPUT ('E', $BM6, '2 /'$);
 65. OUTPUT ('T', 'КООРДИНАТЫ:ВСОБСТВЕННЫХ:ВВЕКТОРОВ', '1/');
 66. FOR $J:=1 \text{ STEP } 1 \text{ UNTIL } I \text{ DO BEGIN FOR } K:=1 \text{ STEP } 1 \text{ UNTIL } KП \text{ DO } BM3[K]:=CB[J,K];$
 67. OUTPUT ('E'. $BM3, '/'$) END ;OUTPUT ('2 /');
 68. IF $KOB=1$ THEN GOTO M4;
 69. OUTPUT ('T', 'ТАБЛИЦА:ВОБЪЕКТОВ:ВВ:ВНОВОЙ:ВСИСТЕМЕ:РКООРДИНАТ', '1/');
 70. FOR $J:=1 \text{ STEP } 1 \text{ UNTIL } KOB \text{ DO BEGIN FOR } K:=1 \text{ STEP } 1 \text{ UNTIL } I \text{ DO BEGIN } BП1:=0; \text{FOR } S:=1$
STEP 1 UNTIL KП DO

71. $ВП1 := ВП1 + СВ[К, S] \times ВМ1[J, S] ; ВМ6[К] := ВП1$ END ;
72. OUTPUT ('E', ВМ6, ' / ') END END ;
73. М4: OUTPUT ('4 / ') END ;
74. ОСНОВНАЯ(1,3, 10^{-3} , 0.95); ОСНОВНАЯ (1,4, 10^{-3} , 0.95) ;
75. ОСНОВНАЯ(1,5, 10^{-5} , 0.99); ОСНОВНАЯ (8,4, 10^{-5} , 0.9) ; END

Длина программы 4000 - 05734

ПРОГРАММА П 2.

"Корреляционное отношение".

Составлена на языке алфа для ЭВМ типа
М-220, М-222.

Бурова Е.И., Карбышев В.Д.

Назначение.

Для изучения парных связей между характеристическими признаками в геологических объектах широко используются статистические меры корреляции [1]. Для оценки линейной корреляционной взаимосвязи между признаками часто применяется при геологических исследованиях коэффициент корреляции, а при изучении между двумя признаками зависимости, отличающейся от линейной, используется корреляционное отношение. Определение значимости отличия исследуемых зависимостей от линейных может производиться с помощью величины Q , приведенной в работе В.Н.Бондаренко [1].

Кроме того, корреляционное отношение позволяет выделять при изучении связей между двумя признаками (x и y) зависимость значений признака y от x и зависимость x от y . Отсюда несомненный интерес представляют возможные зависимости между признаками x и y , когда корреляционное отношение, например, в направлении от признака x к y значительно превышает величину корреляционного отношения в направлении от признака y к x . На основании чего можно делать выводы о возможном влиянии значений одного признака на другой [3].

Отметим одну особенность корреляционного отношения. Вычисление его требует предварительной группировки (разбивки на интервалы) значений признаков, являющейся необходимой операцией для получения условных (групповых) средних. Очень часто решение вопроса о количестве интервалов основано на опыте и интуиции исследователя, без учета количества данных (наблюдений, анализов) и характера распределения значений того или иного признака, что не совсем оправдано и зачастую ведет за собой искажение первоначальной информации, заложенной в материале, т.е. для сгруппированных данных значительно может измениться величина и форма зависимости между признаками, а это может привести в свою очередь к получению неверных результатов и выводов.

Предположим что, после группирования значений признаков, ве-

личина и форма зависимости между признаками не изменилась, если величина коэффициента корреляции до и после группирования осталась постоянной. Тогда использование формулы Стерджесса для определения величины интервала дает удовлетворительные результаты при группировании значений признаков.

С учетом выше сказанного была составлена программа для вычисления корреляционных отношений.

Краткое описание алгоритма 2.

1. Общие сведения.

Программа составлена для подсчета корреляционных отношений с целью изучения зависимости между геологическими характеристиками (признаками), отличающейся от линейной (то есть криволинейной зависимости). Корреляционное отношение между двумя геологическими характеристиками (X, Y) определяется по формулам (1,2) [1].

$$\eta_{x/y} = \frac{S_{xy}}{S_x} \quad (x \text{ от } y) \quad (1), \quad \eta_{y/x} = \frac{S_{yx}}{S_y} \quad (y \text{ от } x) \quad (2)$$

где S_x и S_y - оценки общего стандартного отклонения от признаков X и Y ;

S_{x_i} и S_{y_i} - оценки стандартного отклонения групповых средних, вычисляемые по следующим формулам:

$$S_{x_i} = \sqrt{\frac{1}{N} \sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2}, \quad S_{y_i} = \sqrt{\frac{1}{N} \sum_{i=1}^m n_i (\bar{y}_i - \bar{y})^2},$$

где n_i - число наблюдений в i -том интервале, N - общее число наблюдений, m - число интервалов, \bar{x}_i и \bar{y}_i - средние для характеристик $i^{\text{го}}$ интервала, а \bar{x} и \bar{y} - общие средние для признаков X и Y .

Разбивка на интервалы, операция, необходимая при расчете η , производится по формуле Стерджесса:

$$d = \frac{x_{\max} - x_{\min}}{1 + 3,32 \sqrt[3]{\lg N}},$$

где x_{\max} , x_{\min} - соответственно максимальное и минимальное значение признака x , N - число наблюдений, d - величина интервала.

Инструкция к пользованию.

Программа рассчитана на обработку массива информации объемом $q = 12000$ знаков (в десятичной системе).

Подготовка данных.

Массив информации набивается по столбцам (отдельно каждый признак один за одним) в десятичной системе.

Если число знаков в массиве информации (объем матрицы информации) $q_{10} > 1023$, то массив информации разбивается на группы. Объем каждой группы (число знаков в ней, или количество значений признаков $N \cdot m \leq 1023$ знаков, где N - число наблюдений, m - число признаков).

Подготовка к счету.

Программа (рабочая программа) состоит из 44 перфокарт, из которых 44-я перфокарта является правильной к Σ (контрольная сумма).

После программы ставится информационная карта № 1, на которой пробито число наблюдений (N) с правильной к Σ для этой карты.

Пример:

порядок	N
+ 2	26
+ 2	26

+ к Σ

и информационная карта № 2, на которой пробито число, равное количеству признаков (m) с правильной к Σ для этой карты.

Пример:

порядок	m
+ 2	10
+ 2	10

+ к Σ

После информационных карт ставится массив информации, разбитый с учетом вышесказанного на группы. После каждой такой группы информации ставится к Σ , а в конце-перфокарта, пробитая в командах

7 77 7777 7777 7777

и правильная к Σ (контрольная сумма).

Если массив данных вообще меньше $N \cdot m < 1023$, то после него ставится такая же перфокарта, пробитая в командах, и правильная к Σ

На печать выдаются следующие величины:

- 1) N - число анализов;
- 2) m - число признаков;
- 3) 1-ый фиксированный признак;
- 4) номера и соответствующие им корреляционные отношения последующих признаков;
- 5) 2-ой фиксированный признак;
- 6) номера и корреляционные отношения и т.д.

Контрольный пример.

Исходная таблица.

№	а	б	№	а	б	№	а	б	№	а	б
1	3	0	11	2,8	10,5	21	5,8	5,5	31	9	13
2	3	5	12	2,8	11,5	22	10,8	10,5	32	9	17
3	2,5	1	13	3	11	23	11,5	11	33	9,5	14
4	3,5	4	14	3,5	10	24	8,5	19	34	10,5	10
5	0	4,5	15	3,8	10,5	25	6,5	20	35	11	10
6	1	4	16	4,2	4,5	26	12,5	5	36	11	11
7	1,5	4	17	4,5	5	27	7,2	21	37	11,2	10,5
8	2,4	9,5	18	5	5	28	7,8	19,5	38	11,8	10,5
9	2,4	11,5	19	5,2	3,5	29	8,5	14	39	13,5	4
10	2,5	10	20	5,8	3,5	30	8,5	17	40	14	4,5

где № - число анализов, а и б - признаки.

Исходная информация, подготовленная к счету.

Информационная карта № 1.

π	КОП	A ₁	A ₂	A ₃
+	02	400	-	-
+	02	400	-	-
				кΣ

Информационная карта № 2.

π	КОП	A ₁	A ₂	A ₃
+	01	200	-	-
+	01	200	-	-
				кΣ

Массив исходных данных, перфорирующийся по столбцам.

Первый столбец (а)

π	КОП	A ₁	A ₂	A ₃
+	1	300	-	-
+	1	300	-	-
+	1	250	-	-
+	1	300	-	-
.

+	2	135	-	-
+	2	140	-	-

Второй столбец (б)

\mathcal{K}	КОП	A_1	A_2	A_3
+	0	000	-	-
+	I	500	-	-
+	I	100	-	-
+	I	400	-	-
⋮	⋯	⋯	⋯	⋯
+	I	400	-	-
+	I	450	-	-
⋮				

Результаты счета

$$N = 40$$

$$m = 2$$

1-й фиксированный признак № I

2-й признак; $\eta_1(2) = 0,743$

2-й фиксированный признак № 2

1-й признак $\eta_2(1) = 0,352$

Литература.

1. В.Н.Бондаренко. Статистическое решение некоторых задач геологии. Из-во "Недра", М., 1970, стр.101-111.
2. А.П.Ершов, Г.И.Кожухин, И.В.Поттосин. Руководство к пользованию системой Альфа. ВЦ СОАН СССР, Новосибирск, 1968.
3. О.П.Вышемирская, В.Д.Карбышев. Опыт математической обработки геохимических данных в нефтяной геологии. В печати "Наука".

ТЕКСТ ПРОГРАММЫ П2

```

начало целый p,N,m,l,i,j,k ,*Сигма*ш,*мю*ш, Зонао,швз, Зона, N Зок;
ввод (N,m); Зонао:=1;
начало вещ массив d[1:1024],Y,X[1:N];
вещ minX,maxX,I, *гамма*1,*гамма*2,пх ,*сигма* Y,N*сигма*Y ,*Сигма**эта*1 ,*Сигма**эта*
2,*эта*,
*альфа*,x,y;
целый *Сигма*з;
вещ массив БАР [1 :12,1:1024] ; барабан БАР;
вывод(N,m,ложь);
начало целый гран,i; вещ cfin;
целый чз,чсз,гранк;
чсз:=entier (1023/N); чз:=entier((m-0,5)/чсз)+1;
гранк:=(m-чсз*(чз-1))* N+1;
гран:=N*entier ( 1023 /N)+1; I:=0;
Метзап:i:=i+1; если i>12 то стоп ; d[ ]:=0;
если i=чз то гран:=гранк;
СП 0176(0016,d[1],d [гран],0,0);

```

```

cfin:=d [гран]; d[гран]:=0;
СП0177(1037,d[1],d[1024],0 , БАР[1,1]);
М1:КОД (777,7777,7777,7777);
КОД (015,cfin,М1,0);
КОД(076,0, Метзап,0);
конец ;
к:=0; мвз:= entier(1023/N);
Метк:к:=к+1; *альфа*:= (к-0.5)/мвз; N Збк:=Зонао+ entier (*альфа*);
вывод(к,ЛОЖЬ);
СП0177(1032,d[1],d[1024],0, БАР[N Збк, 1]);
для j:=1,...,1024цикл начало x:=d[j]; СП0002(x,y); d[j]:=y конец ;
l:=к- мвз* entier (* альфа* );
для j:=1,...,N цикл начало p:=j+(1-1)*N; X[j]:=d[p] конец ;
*альфа*:=X[1]; для j:=2,...,N цикл если X[j]<* альфа* то *альфа*:=X[j]; min X:=* альфа* ;
*альфа*:=X[1]; для j:=2,..., N цикл если X[j]>* альфа* то *альфа*:=X[j]; max X:=* альфа* ;
j:=entier(1+3.32*ln(N)/ln(10)+0.5);
I:=(maxX - minX)/j;
Зона:=Зонао; *Сигма*ш:=0; *Сигма*з:=1;

```

если $m < m_{\text{вз}}$ то $*M^*_{\text{м}} := m$ иначе $*M^*_{\text{м}} := m_{\text{вз}}$;

MetЗона:

СП0177(1032, d[1], d[1024], 0, БАР [Зона, 1]);

для $j := 1, \dots, 1024$ цикл начало $x := d[j]$; СП 0002(x, y); $d[j] := y$ конец ;

$i := *Сигма^*_{\text{м}} + 1$; Met i :N* σ *Y:=*Сигма**эта*2:=0; *Сигма**эта*1:=0;

*Сигма** м *:=*Сигма** $\text{м}+1$;

если $i = k$ то начало $i := i + 1$; *Сигма** м *:=*Сигма** $\text{м}+1$; если *Сигма** м * > $m_{\text{м}}$ то на конец ;

для $j := 1, \dots, N$ цикл начало $p := j + (i - 1) * N$; Y[j]:=d[p];

N* σ *Y:=N* σ *Y+Y[j]/N конец ;

для $j := 1, \dots, N$ цикл *Сигма**эта*2:=*Сигма**эта*2+(Y[j]-N* σ *Y) \uparrow 2;

* γ *1:=minX; Met * γ *:=* γ *2:=* γ *1+1; если * γ *2 > maxX то * γ *2:=maxX ;
 $\text{пх} := 0$; * σ *Y:=0;

для $j := 1, \dots, N$ цикл начало если X[j] < * γ *2 and X[j] > * γ *1 то начало
 $\text{пх} := \text{пх} + 1$; * σ *Y:=* σ *Y+Y[j] конец конец ;

если $\text{пх} \neq 0$ то * σ *Y:=* σ *Y/ пх ;

*Сигма**эта*1:=*Сигма**эта*1+ пх *(σ *Y-N* σ *Y) \uparrow 2;

если * γ *2 < maxX то начало * γ *1:=* γ *2; на Met* γ * конец ;

```

*Сигма**эта*1:=*Сигма**эта*1/N ; *Сигма**эта*2:=*Сигма**эта*2/( N-1);
*эта*:=sqrt(*Сигма**эта*1/*Сигма**эта*2);
вывод(*Сигма*м,*эта*,*мю*м,*Сигма*З);
  если *Сигма*м>*мю*м то на endi;
i:=i+1; на Meti ;
endi: если *мю*м=m то на end З;
Зона:=Зона+1; *Сигма*З:=*Сигма*З+1;
если м>мвз * *Сигма*З то *мю*м:=*мю*м+мвз иначе
*мю*м:=*мю*м+(м-(*Сигма*З-1)*мвз); на MetЗона;
endЗ: если к<м то на Metк;
конец конец *

```

ТАБЛИЦА РАСПРЕДЕЛЕНИЯ ПАМЯТИ

181

	начало	конец	длина	примечания
программа	0063	0706	0624	
числовые константы	0707	0736	0030	
формируемые константы	0737	0763	0025	
скалярн	0764	1026	0043	
рабочие ячейки	1027	1035	0007	
массивы				
1	xxxx	xxxx	xxxx	не распределен
2	xxxx	xxxx	xxxx	не распределен
3	1036	3035	2000	
4	020000	027777	010000	
первое рабочее поле ас	начало	конец	длина	
	3036	7303	4246	
рабочее поле ис	начало	конец	длина	
	7304	7500	0174	

адреса программных остановов

0201

ПРОГРАММА АППРОКСИМАЦИИ КРИВЫХ (АК)
А.И. Прокопенко, Ю.Я. Лыков, М.А. Левчук

Программа написана на языке "АЛЬФА" и транслировалась с помощью транслятора ТА-ПМ на ЭВМ типа М-220, М-222.

Назначение

Программа предназначена для аппроксимации кривых линий, образующихся вследствие пересечения поверхности слоистой толщи вертикальной плоскостью и получения параметров радиуса кривизны, центров мгновенных радиусов кривизны. Анализ этих данных позволяет решить ряд конкретных вопросов, таких как миграция свода складки с глубиной, степень асимметрии ее крыльев и т.д.

Краткое описание алгоритма.

Анализируемые кривые линии аппроксимируются показательной функцией $y = a_1 \cdot e^{a_2 x}$ и рассматриваются в системе прямоугольных координат (x, y) , где точка перегиба совпадает с центром и имеет параметры $(0; 0)$. Коэффициенты a_1 и a_2 находятся (определяются) способом наименьших квадратов. Согласно этому методу подходящей теоретической кривой к данной фактической будет та, которая удовлетворяет следующему условию: сумма квадратов отклонений ϵ_i всех ординат фактической кривой от наиболее подходящей теоретической составляет величину минимальную. При этом считается, что отклонения ϵ_i подчиняются нормальному закону распределения. Пусть $y = a_1 \cdot e^{a_2 x}$ (1), логарифмируя это равенство, получаем:

$$\ln y = \ln a_1 + a_2 \cdot \ln e \cdot x;$$

$$\ln y = y'; \quad \ln a_1 = a_0; \quad a_2 \cdot \ln e = a_1 \quad (2)$$

$$\sum_{i=1}^n [y'_i - (a_0 + a_1 x_i)]^2 = \min \quad (3)$$

Далее дифференцируем сумму квадратов отклонений по a_0 и a_1 , и, приравнявая к нулю производные, получим систему уравнений с двумя неизвестными.

$2 \sum_{i=1}^n [y'_i - (a_0 + a_1 x_i)] = 0$; $2 \sum_{i=1}^n [y'_i - (a_0 + a_1 x_i)] \cdot x_i = 0$
 Перепишем их в упрощенном виде, удобном для решения:

$$y'_1 - a_0 - a_1 x_1 + y'_2 - a_0 - a_1 x_2 + \dots + y'_n - a_0 - a_1 x_n = 0$$

$$n a_0 = \underbrace{[y'_1 + \dots + y'_n]}_A - a_1 \underbrace{[x_1 + \dots + x_n]}_B$$

$$y'_1 x_1 - a_0 x_1 - a_1 x_1^2 + y'_2 x_2 - a_0 x_2 - a_1 x_2^2 + \dots + y'_n x_n - a_0 x_n - a_1 x_n^2 = 0$$

$$a_0 [x_1 + \dots + x_n] = \underbrace{[y'_1 x_1 + \dots + y'_n x_n]}_C - a_1 \underbrace{[x_1^2 + \dots + x_n^2]}_D$$

получим:

$$\begin{cases} n a_0 = A - a_1 B & | \times B \\ a_0 \cdot B = C - a_1 D & | \times n \end{cases}$$

отсюда:

$$\left. \begin{aligned} a_1 &= \frac{C \cdot n - A \cdot B}{B \cdot n - B^2} \\ a_0 &= \frac{A}{n} - \frac{B}{n} \cdot \frac{C \cdot n - A \cdot B}{B \cdot n - B^2} \end{aligned} \right\} (4)$$

Затем, подставляя (4) в (2), найдем y ; a_1 ; a_0 .

Далее, подставляя в найденную зависимость данные x , найдем теоретическое y .

После того, как найдено аналитическое выражение кривой по формулам, заданным в явном виде, рассчитываются радиусы кривизны:

$$R = \frac{[1 + (\frac{dy}{dx})^2]^{\frac{3}{2}}}{d^2 y / dx^2}$$

Координаты центра мгновенных радиусов кривизны для каждой пары чисел (x, y) .

$$x_c = x - \frac{\frac{dy}{dx} [1 + (\frac{dy}{dx})^2]}{\frac{d^2 y}{dx^2}}; \quad y_c = y + \frac{1 + (\frac{dy}{dx})^2}{\frac{d^2 y}{dx^2}}$$

Инструкция к пользованию

Данные должны быть представлены в следующем виде:

- 1) массив наблюдаемых значений. x ;
- 2) массив наблюдаемых значений. y ;
- 3) число точек экспериментальной кривой. N ;
- 4) масштабный коэффициент, позволяющий работать с кривыми,

у которых вертикальный масштаб крупнее горизонтального (в

$$L \text{ раз}) - L.$$

Для того, чтобы не происходило остановок при решении из-за переполнения разрядной сетки, вычисление радиусов кривизны производится с предварительной проверкой порядка результатов вычисления.

Число точек исходных данных может быть неограниченным.

Последовательность ввода исходных данных следующая:

- 1) N
- 2) x
- 3) y
- 4) L

Выдача на печать производится в следующем порядке:

- 1) Число точек кривой N
- 2) Массив наблюдаемых x
- 3) Массив наблюдаемых y
- 4) Коэффициент a_0 функции $y = a_0 \cdot e^{a_1 x}$
- 5) Коэффициент a_1 функции $y = a_0 \cdot e^{a_1 x}$
- 6) S - средние координаты центра кривизны для кривой по x .
- 7) $S1$ - средние координаты центра кривизны для кривой по y .
- 8) SS - массив координат центров кривизны для всех точек полученной кривой по x .
- 9) $SS1$ - массив координат центров кривизны для всех точек полученной кривой по y .
- 10) RR - массив радиусов кривизны для всех точек полученной кривой.
- II) - массив полученных по формуле $y = a_0 e^{a_1 x}$ при наблюдаемых x_0 .

Контрольный пример

Ниже предлагается задача при $N = 15$, $L = 1$ и значениями x и y .

Результаты счета приводятся в следующем виде:

N	I	2	3	4	5	6	7	8	9	10	II	I2	I3	I4	I5
X	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0	11,0	12,0	13,0	14,0	15,0
y	14,3	12,6	10,7	9,50	8,20	7,30	6,30	5,40	4,60	3,80	3,20	2,70	2,00	1,50	1,00
$a_0 = 19,63;$ $a_1 = 0,175;$ $S = 24,35;$ $S_1 = 21,47;$															
N	I	2	3	4	5	6	7	8	9	10	II	I2	I3	I4	I5
SS	54,38	41,23	32,28	26,28	22,36	19,89	18,46	17,74	17,54	17,69	18,19	18,68	19,38	20,17	21,03
SS	34,89	29,96	25,95	22,75	20,26	18,39	17,10	16,33	16,07	16,32	17,06	18,34	20,19	22,66	25,83
RR	56,47	42,42	32,62	25,82	21,17	18,08	16,14	15,09	14,77	15,05	15,90	17,30	19,28	21,88	25,16
M	16,46	13,81	11,58	9,71	8,14	6,83	5,73	4,80	4,03	3,38	2,83	2,37	1,99	1,67	1,40

Л И Т Е Р А Т У Р А

1. Руководство к пользованию системы "АЛЬФА". Новосибирск, НГУ, 1974.
2. Гутер Р.С., Овчинский Б.В. Элементы численного анализа и математической обработки результатов опыта. "Наука", М., 1970.

```

НАЧАЛО ВЕЩ  $\bar{A}$ ,  $\bar{B}$ ,  $\bar{C}$ ,  $A_0$ ,  $A_1$ ;
ЦЕЛЫЙ I, N, L;
ВВОД(N); НАЧАЛО
МАССИВ X, Y, YU[1:N];
ВВОД(X, Y, L);
ВЫВОД(N, X, Y, L, ДОЖЬ);
 $\bar{A} := \bar{B} := \bar{C} := D := 0$ ; ДЛЯ I := 1, ..., N ЦИКЛ
НАЧАЛО YU[I] := LN(Y[I]/L);
 $\bar{A} := \bar{A} + YU[I]$ ;  $\bar{B} := \bar{B} + X[I]$ ;
 $\bar{C} := YU[I] \times X[I] + \bar{C}$ ;  $\bar{D} := \bar{D} + X[I] \times X[I]$ ; КОНЕЦ;
 $A_1 := (\bar{C} \times N - \bar{A} \times \bar{B}) / (\bar{D} \times N - \bar{B} | 2)$ ;
 $A_0 := \bar{A} / N - \bar{B} \times (\bar{C} \times N - \bar{A} \times \bar{B}) / (\bar{D} \times N | 2 - N \times \bar{B} | 2)$ ;
 $A_0 := \text{EXP}(A_0)$ ;
НАЧАЛО ЦЕЛЫЙ K; ВЕЩ S, S1,  $\bar{R}$ ;
МАССИВ SS, SS1,  $\bar{R}$   $\bar{R}[1:N]$ ;
 $\bar{R} := S := S1 := 0$ ;
ДЛЯ K := 1, ..., N ЦИКЛ НАЧАЛО SS[K] := X[K] -  $A_0 \times A_1 \times \text{EXP}(A_1 \times X[K])$ 
 $\times (1 + (A_0 \times A_1 \times \text{EXP}(A_1 \times X[K])) | 2)$ 
 $/ (A_0 \times A_1 | 2 \times \text{EXP}(A_1 \times X[K]))$ ;

```

```

SS1[K]:= AO*EXP(A1*X[K])
+(1+(AO*A1*EXP(A1*X[K]))|2,
/(AO*A1|2*EXP(A1*X[K]));
S:= S+SS[K]/N;    S1:= S1+SS1[K]/N;
R R[K]:= SQRT((1+(AO*A1*EXP(A1*X[K]))|2)|3)/ABS(AO*A1|2*EXP(A1*X[K]));
R:= R+ R R[K]/N;
КОНЕЦ;    НАЧАЛО ЦЕЛЫЙ I;    МАССИВ M[1:N];
ДЛЯ I:=, ..., N ЦИКЛ
M[I]:= L*AO*EXP(A1*X[I]);
ВЫВОД(AO, A1, ДОЖЬ, S, S1, ДОЖЬ, SS, ДОЖЬ, SS1, ДОЖЬ, R R, ДОЖЬ, M)
КОНЕЦ    КОНЕЦ    КОНЕЦ    КОНЕЦ *

```

С О Д Е Р Ж А Н И Е

ВВЕДЕНИЕ	3
ЧАСТЬ I. ПРОГРАММЫ МЕТОДА СОГЛАСОВАННЫХ ОЦЕНОК	6
А.Н.Дмитриев, В.О.Красавчиков	
1. Кандыба В.Н. Программа П1 "Качели для бинарных таблиц"	14
2. Кандыба В.Н. Программа П2 "Расчет коэффициентов"	20
3. Смертин Е.А. Программа П3 "Симметрия"	27
ЧАСТЬ II. ПРОГРАММЫ "ТЕСТОВЫХ ОЦЕНОК" (ТЕСТЫ,ТЕСТОРЫ).	36
А.Н.Дмитриев, Т.Л.Слущкая, Е.А.Смертин,Т.И.Штатнова	
4. Слущкая Т.Л. Программа П1 "Вычисление информационных весов признаков" для ЭВМ М-220	46
5. Слущкая Т.Л. Программа П2 "Вычисление информационных весов признаков" для ЭВМ БЭСМ-6	54
6. Слущкая Т.Л. программа П3 "Вычисление тесторной информационной меры признаков"	66
7. Слущкая Т.Л. Программа П4 "Диагностика объектов голосованием по тестам"	77
8. Слущкая Т.Л. Программа П5 "Диагностика объектов голосованием по тесторам"	90
9. Слущкая Т.Л. Программа П6 "Вычисление тестовых и тесторных параметров таблиц с учетом длин тупиковых тестов и тесторов"	102
10. Смертин Е.А. Программа П7 "Состав-1"	114
ЧАСТЬ III. ПРОГРАММЫ МЕТОДА СУММАРНОГО УЧЕТА МЕР ПРИУРОЧЕННОСТИ и СОГЛАСОВАНИЯ	122
Бабищ В.В., В.О.Красавчиков, Г.С.Федосеев	
11. Кандыба В.Н. Программа П1 "Оценка информативности признаков"	127
12. Кандыба В.Н. Программа П2 "Оценка информативности признаков"	131

13. Бабич В.В. Программа ПЗ "Подсчет строчечных нагрузок по заданной информативности для бинарных таблиц".	135
14. Бабич в.В. Программа П4 "Оптимальное бинарное кодирование признаков".	143
15. Соколов А.Д. Программа П5 "Оптимальное бинарное кодирование признаков, подсчет строчечных нагрузок и минимизация".	151
ЧАСТЬ IV. ПРОГРАММЫ ДЛЯ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ГЕОЛОГИЧЕСКОЙ ИНФОРМАЦИИ	
16. Бурова Е.И., Карбышев В.Д. Программа П1 "Главные компоненты"	159
17. Бурова Е.И., Карбышев В.Д. Программа П2 "Корреляционные отношения".	172
18. Прокопенко А.П., Лыков Ю.Я., Лычук И.А. Программа аппроксимации кривых (АК).	182

Технический редактор *Л. А. Панина*

Подписано к печати 2.УП.1975г. МН 03043
Бумага 60×84/16. Печ.л. 12.0 Уч.-изд. л. 10.0
Тираж 600 **Заказ 239** Цена 70 коп.

Институт геологии и геофизики СО АН СССР
Новосибирск, 90. Ротапринт.