АНАЛИЗ ДАННЫХ В ЭКСПЕРТНЫХ СИСТЕМАХ
(Вычислительные системы)
1986 год Выпуск 117

УДК 621.391:512.2

ПОСТРОЕНИЕ ДЕРЕВА РАЗБИЕНИЙ В ЗАДАЧЕ ГРУППИРОВКИ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ ЛОГИЧЕСКИХ ФУНКЦИЙ

Г.С.Лбов, Т.М.Пестунова

Введение

Данная работа продолжает исследования, связанные с разработкой методов и алгоритмов анализа данных на основе использования логических функций [1,2]. Эти методы позволяют учесть следующие часто встречающиеся на практике особенности таблиц данных: многомерность, разнотипность, наличие пропусков, отсутствие априорных сведений, позволяющих делать какие-либо предположения о модели изучаемого явления и др.

К числу наиболее распространенных задач анализа данных относится таксономия (группировка). Для ее решения также могут быть успешно применены методы, основанные на логических функциях.В работе [2] рассматривался алгоритм, реализующий локально-оптимальную процедуру, позволяющую последовательно выделять группы объектов,имеющих сходные значения по большинству признаков, и получать описания этих групп в виде логических закономерностей. Этот алгоритм можно также использовать для выделения закономерностей, характерных для всей эмпирической таблицы, в виде набора конъюнкций с пересекающимися множествами истинности.

Для решения задач распознавания с использованием логических функций разработаны алгоритмы построения решающих правил в виде дерева, основанные на простых оптимизационных процедурах и позволяющие при распознавании объекта минимизировать число проверяемых на нем элементарных высказываний [1]. В данной статье предлагается алгоритм построения дерева разбиений для решения задачи группировки объектов. Заметим, что свойство иерархичности, которым об-

ладает группировка в виде дерева, часто оказывается полезным для выяснения структуры множества данных, так как дает возможность получать на нем разбиения различной степени генерализации.

§I. Определения

Пусть имеется некоторое множество объектов $A=\{a_1,\ldots,a_i,\ldots,a_N\}$, выбранное из генеральной совокупности Γ и характеризуемое набором признаков $X=\{X_1,\ldots X_j,\ldots X_n\}$. Множеству A сопоставляется набор реализаций $t=\{x_1,\ldots,x_i,\ldots,x_N\}$, где $x_i = (x_i^1,\ldots,x_i^j,\ldots,x_i^n)$, а x_i^j – результат измерения X_j у a_i. Получаем эмпирическую таблицу $v=\{x_i^j\}$, $i=\overline{1,N}$; $j=\overline{1,n}$, по которой некоторым образом строится признаковое пространство $D=D(v)$. Пусть R_t и R_A – множества всевозможных разбиений на группы множеств t и A. Каждому $r_t\in R_t$, $r_t=\{t_1,\ldots,t_\omega,\ldots,t_k\}$, $t_\omega\neq\emptyset$ (а ему соответствует некоторое $r_A\in R_A$, $r_A=\{A_1,\ldots A_\omega,\ldots,A_k\}$, $A_\omega\neq\emptyset$), сопоставляется набор таксонов $\alpha=(T_1,\ldots T_\omega,\ldots T_k)$ в пространстве D, удовлетворяющий условиям:

$$t_\omega\subseteq T_\omega,\quad \omega=\overline{1,k},$$

$$T_i\cap T_j=\emptyset\quad \forall i\neq j,$$

$$\bigcup_{\omega=1}^{k} T_\omega\subseteq D.$$

Задача состоит в нахождении набора таксонов α и соответствующих ему разбиений r_t и r_A, минимизирующих некоторый критерий качества F. Для описания таксонов в том или ином методе группировки используется некоторый класс функций Φ.

В разработанном методе группировки используется класс логических функций. Ниже дается описание этого класса.

Признаковое пространство D рассматривается в виде $D=\prod_{j=1}^{n} D_j$, где D_j – множество значений признака X_j, которое определяется по таблице v с учетом типа этого признака. Для порядковых и номинальных признаков D_j включает в себя все несовпадающие их значения на эмпирической таблице, для количественного – замкнутый интервал, ограниченный минимальным и максимальным его значениями.

На D_j определяются множества d_j (также в зависимости от типа признака):

d_j есть значение или произвольный набор значений для номинального X_j;

d_j есть значение либо набор соседних значений для порядкового X_j;

d_j есть замкнутый интервал на D_j для количественного X_j типа $[x_i^j, x_v^j]$, где x_i^j и x_v^j – значения признака X_j для множества реализаций t $(x_i^j \le x_v^j)$.

Для произвольной реализации $x \in t$ и множества $d_j \subseteq D_j$ элементарной высказывательной функцией называется двуместный предикат со значениями:

$$I(x,d_j) = \begin{cases} 1, & \text{если } X_j(x) \in d_j; \\ 0, & \text{если } X_j(x) \notin d_j \text{ или не измерено,} \end{cases}$$

где $X_j(x)$ – значение j-й компоненты реализации x. Таксоны описываются с помощью конъюнкций $S(x,d') = \&_{m=1}^{M} I(x,d_{j_m})$, определенных на некотором подмножестве признаков $X' \subseteq X$ с соответствующим признаковым пространством $D' \subseteq D$ размерности $M \le n$; $d' \subseteq D'$ и $d' = \prod_{m=1}^{M} d_{j_m}$. При этом d' – множество истинности S, которое является таксоном для группы реализаций $\mathfrak{T}$, таких что $\forall x \in \mathfrak{T}$ $S(x,d') = 1$. Соответственно определяется и группа объектов.

Считается, что значения признаков из $X \setminus X'$ могут быть произвольными для объектов из данной группы.

§2. Критерий группировки объектов

Критерии, применяемые для оценки качества таксонов и разбиений, основаны на двух функциях: $P(S)$ – вероятность истинности конъюнкции S при условии, что реализация x выбирается из равномерного распределения вероятностей в n-мерном пространстве признаков, $\nu(S)$ – фактическая частота истинности S на данной эмпирической таблице [2].

При реализации алгоритмов группировки используются два критерия:

– для оценки качества отдельно взятого таксона T, определяемого как множество точек n-мерного пространства, для которых истинна конъюнкция S, используется критерий: $f(s) = \nu(S) - P(S)$;

- для оценки качества расположения пары таксонов T_1 и T_2, соответствующих конъюнкциям S_1 и S_2, используется другой критерий: $\Delta(S_1,S_2) = (P(S_1) + P(S_2))/P(S_{12})$, где S_{12} определяется из условий: I) $T_{12} \supseteq T_1 \cup T_2$ и 2) $f(S_{12}) = \max f(S)$ среди всех S, удовлетворяющих первому условию.

При этом $f(S) \to \max$, $\Delta(S_1,S_2) \to \min$. В случае разбиения фиксированного множества объектов на два таксона разбиение, которому соответствует $\min \Delta(S_1,S_2)$, дает максимальное значение $f(S_1) + f(S_2)$, что эквивалентно в случае количественных признаков разбиению с минимизацией суммарного объема таксонов, описывающих выделенные группы объектов. Более подробно свойства критериев $f(S)$ и $\Delta(S_1,S_2)$ и их взаимосвязь изложены в работе [3].

Основная идея алгоритма построения разбиения на таксоны в виде бинарного дерева состоит в увеличении на каждом шаге числа таксонов на единицу путем разделения одного из имеющихся к данному моменту таксонов на два в соответствии с некоторым критерием. На первом шаге делится единственный таксон, содержащий все реализации эмпирической таблицы. Критерий деления основан на использовании функционала $\Delta(S_1,S_2)$. Поскольку перебор на каждом шаге осуществляется по всем имеющимся таксонам и по всем возможным разбиениям каждого таксона в классе логических функций, то зависимость в нем существует уже от трех переменных*):

$$\Delta(S,S_1,S_2) = \frac{P(S_1) + P(S_2)}{P(S)}$$

при условии $t_{S_1} \cup t_{S_2} = t_S$.

Здесь конъюнкция S определяет разделяемый таксон, а S_1 и S_2 - таксоны, получающиеся в результате деления S t_S, t_{S_1}, t_{S_2} - множества реализаций, на которых истинны конъюнкции S, S_1 и S_2 соответственно.

Недостатком разбиений, получаемых при минимизации $\Delta(S_1,S_2)$, является тенденция к разбиению малочисленных групп реализаций на еще более мелкие. Заметим, что еще сильнее это свойство выражено у критерия $\Delta'(S,S_1,S_2) = P(S_1) + P(S_2) - P(S)$, при минимизации которого получается итоговая группировка с минимальным суммарным

*) Построенное путем такого последовательного деления таксонов дерево в дальнейшем будем называть деревом разбиений.

"объемом" таксонов. Поэтому в разработанном алгоритме осуществляется минимизация критерия

$$P(S,S_1,S_2) = \Delta(S,S_1,S_2) = \frac{1}{g(N_S \cdot v(S))},$$

где g – возрастающая монотонная функция, N_S – число реализаций множества t_S. В этом критерии добавлена зависимость от числа реализаций в разделяемом таксоне и преимущество отдается делению крупных таксонов.

Экспериментальные исследования на тестовых задачах и реальных данных показали, что можно ограничиться линейной зависимостью $g(x) = x$. Близкие результаты получаются при использовании функций, возрастающих медленнее: $g(x) = \sqrt{x}$ и $g(x) = \ln x$. Функции возрастающие значительно быстрее, чем $g(x) = x$ (например, $g(x) = x^2$) использовать нежелательно, так как столь сильные зависимости приводят к искажениям и неудовлетворительным результатам.

Если в конъюнкцию S входит элементарная высказывательная функция по количественному признаку на основе вырожденного множества d_j (т.е. имеющего вид $x_0 \le X_j(x) \le x_0$), то вероятность истинности такой конъюнкции $P(S) = 0$ вне зависимости от значений остальных признаков. В таблицах данных совпадение значений количественных признаков встречается на практике довольно часто. Использование для таких таблиц критериев качества разбиений, использующих величину $P(S)$, могло бы привести к тому, что количественные признаки, по которым имеется совпадение значений, получили бы значительно больший вес по сравнению с остальными, а реализации, имеющие совпадение значений хотя бы по одному количественному признаку, во многих случаях выделялись бы в определенный таксон, независимо от значений других признаков.

Однако, работая с таблицами данных, следует помнить, что совпадение значений признаков, измеренных в непрерывных шкалах, зачастую неизбежно, в частности, в силу того, что все измерения фиксируются округленно с некоторой точностью, т.е. фактически осуществляется дискретизация непрерывной шкалы. Отсюда следует, что совпадение значений признаков у объектов не эквивалентно совпадению их табличных значений. Вероятность последнего события безусловно выше (отлична от нуля).

Поскольку при решении задач множества d_j формируются на основе табличных значений признаков, то с учетом вышесказанного можно считать вероятность ε истинности элементарной высказывательной функции с вырожденным множеством d_j по количественному признаку положительной: $\varepsilon > 0$. Другими словами, вырожденный интервал, соответствующий изолированному табличному значению, заменяется невырожденной длиной $\varepsilon \cdot |D_j|$, где $|D_j| = x^j_{max} - x^j_{min}$.

Выбор $\varepsilon > 0$ в соответствии с точностью измерения данных в обрабатываемой таблице не препятствует выделению изолированных во всем пространстве признаков таких таксонов, реализации в которых имеют совпадающие или близкие значения по большинству признаков, в том числе при наличии небольшого числа "шумовых" признаков для данного таксона.

Практически ε выбирается исходя из минимальных расстояний между соседними несовпадающими значениями количественных признаков. Пусть $\varepsilon_j = \dfrac{\min|x^j_i - x^j_l|}{|D_j|}$; $i,l = \overline{1,N}$. Тогда для каждого признака берется свое значение $\varepsilon = \varepsilon_j$. На практике можно воспользоваться одним значением ε для всех признаков, равным минимальному или среднему значению ε_j по всем количественным признакам.

§3. Алгоритм построения дерева разбиений

Алгоритм построения дерева разбиений для решения задачи группировки состоит из двух этапов – предварительного и основного.

В ходе предварительного этапа осуществляется формирование градаций для каждого признака. Для номинальных и порядковых признаков в массив градаций заносятся все несовпадающие их значения на эмпирической таблице. На каждом количественном признаке градации расставляются таким образом, чтобы выделить плотные группы значений этого признака, встречающиеся в таблице, отделенные друг от друга относительно большими промежутками. Эта процедура осуществляется по следующей схеме.

1. Выделяются все несовпадающие значения признака X_j , полученный массив упорядочивается по возрастанию.

2. Вычисляются промежутки между всеми парами соседних значений, массив упорядочивается по убыванию.

3. Полагается $l = [L/2] - 1$, где L – задаваемое заранее максимальное число градаций (обычно оно несколько больше числа таксонов).

4. Выделяются первые l элементов массива, сформированного в п.2.

5. В массив градаций заносятся значения признака X_j, соответствующие элементам, выделенным в п.4.

6. Массив градаций упорядочивается по возрастанию.

7. Удаляются повторяющиеся значения градаций.

8. Добавляются минимальное и максимальное значения X_j на таблице в качестве соответственно первого и последнего элементов массива градаций (если они не были выделены в п.4).

9. Если в результате полученное число градаций значительно меньше L, то $l := l+2$ и переход на п.4. Иначе - на п. I0.

I0. Конец.

Действия пп.I-I0 осуществляются для каждого количественного признака. Для каждого порядкового или номинального признака производятся действия п.I.

Элементы массива градаций, полученного на предварительном этапе алгоритма, используются в ходе основного этапа для формирования множеств d_j, на основе которых строятся элементарная высказывательная функция и конъюнкции-таксоны.

Основной этап непосредственно связан с построением дерева разбиений на множестве реализаций и формированием набора конъюнкций-таксонов, описывающих выделенные группы.

I. Присваивание начальных значений:

N - число объектов,

n - число признаков,

N_{max} - максимальное число таксонов, которое нужно выделить,

n_{max} - число таксонов, выделенных к данному шагу (в начале алгоритма $n_{max} = 1$),

M - массив номеров объектов, записанных в соответствии с их принадлежностью к таксонам. Перед началом алгоритма $M(i) = i \quad \forall i = \overline{1,N}$.

2. $i := 0$ - номер разделяемого таксона.

3. $i := i+1$. Если $i > n_{max}$, то переход на п.I2, иначе выделяется $M_0 \subseteq M$ - множество реализаций, принадлежащих i-му таксону (в начале алгоритма $M_0 = M$).

4. $j := 0$ - номер активного признака (т.е. признака, на основе которого будет строится очередное разбиение),

$F_{опт_i} = \infty$ - начальное значение критерия (перед построением всевозможных разбиений i-го таксона).

5. $j = j+1$. Если $j > n$, то переход на п.3, иначе - на п.7.

6. Если по признаку X_j имеется пропуск среди объектов из M_0, то переход на п.5, иначе - на п.7.

7. Выбирается очередная пара множеств d_1 и d_2 по признаку X_j на объектах из M_0, такая что $d_1 \cap d_2 = \emptyset$, а объединение этих множеств $d_1 \cup d_2$ содержит все значения X_j на объектах из M_0. Если таких d_1 и d_2 выбрать нельзя, то переход на п.5, иначе - на п.8.

8. Строится разбиение объектов из M_0 в соответствии с принадлежностью их значений по X_j к d_1 и d_2. В итоге получаются множества M_1 и M_2, такие что $M_1 \cup M_2 = M_0$ и $M_1 \cap M_2 = \emptyset$.

9. Вычисляются конъюнкции S_1 и S_2 в многомерном пространстве, определяющие таксоны для множеств M_1 и M_2.

10. Вычисляется значение критерия деления F для разбиения множества M на M_1 и M_2, описываемого конъюнкциями S_1 и S_2.

11. Если $F_{опт_i} \le F$, то переход на п.7, иначе $F_{опт_i} = F$ и переход на п.7.

12. Путем сравнения $F_{опт_i}$ $\forall i = \overline{1, n_{max}}$ выделяется i^* такое, что $F_{опт_{i^*}} = \min\limits_{i=\overline{1,n_{max}}} F_{опт_i}$.

13. В массиве M заменяется часть, соответствующая таксону с номером i^*, на переупорядоченный массив M_0 сообразно полученному оптимальному разбиению массива.

В остальных массивах, хранящих информацию о построенных таксонах, удаляются части, соответствующие таксону с номером i^*, и добавляется информация о вновь полученных таксонах.

В массивы, хранящие информацию о структуре дерева, заносятся данные о разбиении таксона i^*.

14. $n_{max} := n_{max} + 1$.

15. Если $n_{max} < N_{max}$, то переход на п.3, иначе - на п.16.

16. Конец.

§4. Проверка эффективности алгоритма на тестовых примерах

Исследование алгоритма проводилось на ряде тестовых задач. Ниже приводятся результаты решения двух из них. В первом приведенном примере использовались разнотипные данные (таблица та же, что и в [2]). Во втором взяты данные по двум количественным признакам,

что позволяет изобразить их и результаты решения наглядно на плоскости. Эти данные использовались ранее для исследования алгоритма "Форэль" [4].

Для удобства описания результатов введем следующие определения.

I. Шаг построения дерева - осуществление последовательности действий п.3-I4 основного этапа алгоритма, приводящих к разделению одного из имеющихся таксонов на два.

2. Узел дерева - множество, являющееся таксоном после некоторого шага построения дерева.

3. Активный признак (на некотором шаге построения дерева) - признак, на основе которого осуществлялось разбиение на этом шаге.

4. Пассивные признаки - признаки, участвующие в описании таксона, за исключением активного.

5. Несущественные признаки (для данного таксона) - признаки, для которых соответствующие им элементарные высказывательные функции в конъюнкции, описывающей таксон, имеют простейшие множества d_j, такие что $d_j = D_j$.

Эмпирическая таблица с разнотипными признаками представлена в табл. I.

Для описания 30 объектов используются 7 признаков, из которых I,3,7 - количественные, 2 - порядковый, 4,5,6 - номинальные. Значение параметра ε выбиралось одинаковым для всех количественных признаков: $\varepsilon = \varepsilon_{\text{средн}} \approx 0{,}01$. Код пропуска равен -I.0.

На первом шаге построения дерева на основе активного признака I осуществилось разделение всего множества данных (узел I) на два таксона (узлы 2 и 3), определяемых конъюнкциями $S_1^{(1)}(x, T_1^{(1)}) = I$ и $S_2^{(1)}(x, T_2^{(1)}) = I$ соответственно. Конъюнкции эти имеют следующий вид:

$$S_1^{(1)}(x, T_1^{(1)}) = \mathop{\&}_{m=1}^{5} I(x, d_{j_m}^{(1)}) = 1,$$

где

$$T_1^{(1)} = \prod_{m=1}^{5} d_{j_m}^{(1)} ;$$

$$j_1 = 1, \quad d_{j_1}^{(1)} = [1.10, 5.15];$$

$$j_2 = 2, \quad d_{j_2}^{(1)} = \{3,4,5,6,8,9\};$$

$$j_3 = 5, \quad d_{j_3}^{(1)} = \{3,4\};$$

$$j_4 = 6, \quad d_{j_4}^{(1)} = \{1,2,4\};$$

$$j_5 = 7, \quad d_{j_5}^{(1)} = [38.0, 60.0]$$

(признаки X_3 и X_5 для такого таксона являются несущественными "шумовыми");

$$s_2^{(1)}(x, T_2^{(1)}) = \&_{i=1}^{7} I(x, d_i'^{(1)}),$$

где

$$T_2^{(1)} = \prod_{i=1}^{7} d_i'^{(1)};$$

$$d_1'^{(1)} = [8.90,10.10]; \; d_2'^{(1)} = \{1,2\}; \; d_3'^{(1)} = [50.0,57.0];$$

$$d_4'^{(1)} = \{0\}; \; d_5'^{(1)} = \{5\}; \; d_6'^{(1)} = \{1,3\}; \; d_7'^{(1)} = [80.0,86.0]$$

("шумовых" признаков нет).

Значения критерия деления $F^{(1)} = 0{,}00258$, критериев качества $f(s_1^{(1)}) = 0{,}555990$, $f(s_2^{(1)}) = 0{,}366589$.

На втором шаге на основе активного признака I произошло разделение узла 2 (таксон T_1^1) и образовались узлы 3 и 4, описываемые конъюнкциями:

$$s_1^{(2)}(x, T_1^2) = \&_{m=1}^{6} I(x, d_{j_m}^{(2)}) = 1 \quad \text{и} \quad s_2^{(2)}(x, T_2^{(2)}) = \&_{m=1}^{5} I(x, d_{i_m}'^{(1)}) = 1,$$

где

$$T_1^{(2)} = \prod_{m=1}^{6} d_{j_m}^{(2)}, \quad T_2^{(2)} = \prod_{m=1}^{5} d_{i_m}'^{(2)}.$$

Множества d_j имеют вид:

$j_1 = 1, \quad d_{j_1}^{(2)} = [1.10, 1.21];$

$j_2 = 2, \quad d_{j_2}^{(2)} = \{8, 9\};$

$j_3 = 3, \quad d_{j_3}^{(2)} = [65.0, 70.0];$

$j_4 = 5, \quad d_{j_4}^{(2)} = \{4\};$

$j_5 = 6, \quad d_{j_5}^{(2)} = \{1, 2\};$

$j_6 = 7, \quad d_{j_6}^{(2)} = [50.0, 60.0];$

$i_1 = 1, \quad d_{i_1}^{\prime(2)} = [4.80, 5.15];$

$i_2 = 2, \quad d_{i_2}^{\prime(2)} = \{3, 4, 5, 6\};$

$i_3 = 5, \quad d_{i_3}^{\prime(2)} = \{3\};$

$i_4 = 6, \quad d_{i_4}^{\prime(2)} = \{4\};$

$i_5 = 7, \quad d_{i_5}^{\prime(2)} = [38.0, 47.0]$

Значения критериев $F^{(2)} = 0{,}00022$, $f(s_1^{(2)}) = 0{,}233317$, $f(s_2^{(2)}) = 0{,}399696$.

На третьем шаге на основе активного булева признака 4 разделился узел 4.

Полученное разбиение характеризуется следующими значениями критерия: $F^{(3)} = 0{,}03078$; $f(s_1^{(3)}) = 0{,}066666$; $f(s_2^{(3)}) = 0{,}166666$.

При этом произошло значительное увеличение значения критерия деления и уменьшение значений критериев качества. Это позволяет сделать вывод о целесообразности остановки после выполнения двух шагов построения дерева.

Таким образом, итоговое разбиение определяется конъюнкциями $S_1 = S_2^{(1)}$, $S_2 = S_1^{(2)}$, $S_3 = S_2^{(2)}$ с множествами истинности $T_1 = T_2^{(1)}$,

$T_2 = T_1^{(2)}$, $T_3 = T_2^{(2)}$. Множество объектов разбилось соответственно на три группы A_1, A_2, A_3, имеющие следующий состав (перечислены номера объектов, попадающих в каждую из групп):

$$A_1 = \{8,9,10,11,12,13,14,15,16,17,18\},$$
$$A_2 = \{1,2,3,4,5,6,7\},$$
$$A_3 = \{19,20,21,22,23,24,25,26,27,28,29\}.$$

Данные второй тестовой задачи приведены в табл.2. Число объектов равно 20, число признаков равно 2 (оба количественные), $\varepsilon = 0{,}09$. Структура дерева разбиений после 4-х шагов построения дерева представлена в табл.3. Динамика изменения значений критериев свидетельствует о нецелесообразности дальнейшего разбиения.

Итоговое разбиение определяется конъюнкциями:

$$S_1(x,T_1) = \mathop{\&}_{j=1}^{2} I(x,d_j^{(1)}) = 1, \quad T_1 = \prod_{j=1}^{2} d_j^{(1)},$$

$$d_1^{(1)} = [7,10], \; d_2^{(1)} = [12,12];$$

$$S_2(x,T_2) = \mathop{\&}_{j=1}^{2} I(x,d_j^{(2)}) = 1, \quad T_2 = \prod_{j=1}^{2} d_j^{(2)},$$

$$d_1^{(2)} = [3,6], \; d_2^{(2)} = [6,9];$$

$$S_3(x,T_3) = \mathop{\&}_{j=1}^{2} I(x,d_j^{(3)}) = 1, \quad T_3 = \prod_{j=1}^{2} d_j^{(3)},$$

$$d_1^{(3)} = [10,11], d_2^{(3)} = [7,9];$$

$$S_4(x,T_4) = \mathop{\&}_{j=1}^{2} I(x,d_j^{(4)}) = 1, \quad T_4 = \prod_{j=1}^{2} d_j^{(4)},$$

$$d_1^{(4)} = [1,2], \; d_2^{(4)} = [1,2];$$

$$S_5(x,T_5) = \mathop{\&}_{j=1}^{2} I(x,d_j^{(5)}) = 1, \quad T_5 = \prod_{j=1}^{2} d_j^{(5)},$$

$$d_1^{(5)} = [6,8], \; d_2^{(5)} = [1,3].$$

Таблица I

Объекты	Признаки						
	I	2	3	4	5	6	7
I	I.I0	9	67.0	0	4	I	50.0
2	I.20	8	70.0	0	4	I	55.0
3	I.I7	-I	66.0	I	4	2	52.0
4	I.2I	8	67.5	I	4	2	57.0
5	I.I2	9	65.0	I	4	2	60.0
6	I.20	9	68.0	I	4	2	60.0
7	I.I6	9	5I.4	I	4	2	52.0
8	I0.00	2	50.5	0	5	3	82.0
9	I0.I0	2	5I.0	0	5	3	85.0
I0	9.30	I	55.0	0	5	I	-I.0
II	9.35	I	56.0	0	5	3	84.0
I2	I0.05	I	57.5	0	5	3	8I.0
I3	9.80	I	56.5	0	5	3	80.0
I4	9.38	2	55.0	0	5	3	85.0
I5	I0.03	2	55.5	0	5	I	83.0
I6	8.90	2	57.0	0	5	I	86.0
I7	8.97	2	-I.0	0	5	I	83.0
I8	9.0I	I	50.0	0	5	I	8I.0
I9	4.97	3	43.0	0	3	4	40.0
20	5.I5	3	40.5	I	3	4	4I.0
2I	4.80	6	70.0	I	3	4	4I.0
22	4.80	4	73.5	I	3	4	47.0
23	4.90	6	40.0	0	3	4	45.0
24	4.85	5	65.0	I	3	4	42.0
25	5.05	4	55.0	I	3	4	-I.0
26	5.07	4	57.0	I	3	4	38.0
27	5.00	3	40.0	I	3	4	39.0
28	5.I0	5	42.0	I	3	4	43.0
29	4.83	4	-I.0	I	3	4	40.0
30	4.82	3	5I.0	0	3	4	40.5

Таблица 2

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

Таблица 3

Структура дерева разбиений

Номер шага	Номер узла делящегося	Номера узлов образованных	Номер признака активного	Тип признака активного	Разбиение по активному признаку		Число объектов в новых узлах	Значение критерия
I	I	2	2	Количественн	I.0000	3.0000	7	0.028I8
		3			6.0000	I2.0000	I3	
2	3	4	2	Количественн	6.0000	9.0000	9	0.04III
		5			I2.0000	I2.0000	4	
3	4	6	I	Количественн	3.0000	6.0000	6	0.05093
		7			I0.0000	II.0000	3	
4	2	8	I	Количественн	I.0000	2.0000	3	0.05I02
		9			6.0000	8.0000	4	

Соответствующие группы объектов:

$A_1 = \{17, 18, 19, 20\}$,
$A_2 = \{3, 10, 11, 12, 13, 14\}$,
$A_3 = \{2, 15, 16\}$,
$A_4 = \{4, 6, 7\}$,
$A_5 = \{1, 5, 8, 9\}$.

Литература

1. ЛБОВ Г.С. Методы обработки разнотипных экспериментальных данных. - Новосибирск: Наука, 1981.

2. ЛБОВ Г.С., ПЕСТУНОВА Т.М. Группировка объектов в пространстве разнотипных признаков. -В кн.: Анализ нечисловой информации в социологических исследованиях /Под ред. Андреенкова В.Г., Орлова А.И., Толстовой Ю.Н. -М., 1985, с.141-149.

3. ПЕСТУНОВА Т.М. О свойствах одного метода группировки, основанного на логических функциях. -В кн.: У школа-семинар по непараметрическим и робастным методам статистики в кибернетике, ч.II. Томск, 1985, с.298-306.

4. ЗАГОРУЙКО Н.Г., ЁЛКИНА В.Н. Программа таксономии ФОРЭЛЬ. Пакет прикладных программ ОТЭКС. Версия 3.0. - Новосибирск, 1980.

Поступила в ред.-изд.отд.
15 сентября 1986 года