

УДК 681.3

С.С. Бучик, ад'юнкт  
Б.М. Герасимов, д.т.н., проф.

Житомирський військовий інститут радіоелектроніки ім. С.П. Корольова

## НЕЧІТКА КЛАСТЕРИЗАЦІЯ ДАНИХ НА ОСНОВІ ПРИНЦИПУ САМООРГАНІЗАЦІЇ

У статті розглядається задача поділу об'єктів, що характеризуються кількісними, отриманими експериментально, і якісними, котрі надані експертами і формалізовані математичним апаратом теорії нечітких множин ознаками, на групи (кластери).

Згідно з [1] кластер-аналіз – це спосіб угруповання багатовимірних об'єктів, заснований на представленні результатів окремих спостережень точками придатного геометричного простору з наступним виділенням груп як “згустків” цих точок.

Таким чином, основна ціль аналізу – виділити у вихідних багатовимірних даних такі однорідні підмножини, щоб об'єкти усередині груп були схожі один на одного, а об'єкти з різних груп – не схожі. Під “подібністю” розуміється близькість об'єктів у багатовимірному просторі ознак, і тоді задача зводиться до виділення в цьому просторі природних скупчень (“грон”) об'єктів, що і вважаються однорідними групами.

Як зазначено в [2], неформально проблему розбивки вибіркової множини на класи можна сформулювати в такий спосіб: “Згрупувати точки вибіркової множини в підмножини (так звані кластери) так, щоб подібні точки відносилися до тієї самої підмножини, а неподібні – до різних підмножин”.

Використання методів кластерного аналізу при розв'язанні широкого кола задач дозволяє адекватно описувати моделі предметних областей реального світу. Кластерний аналіз дозволяє у великих наборах даних виявляти ті структури, що найбільш повно описують поведінку реальних об'єктів.

У статті розглядається задача поділу множини  $S$  об'єктів, кожен  $i$ -й елемент якого характеризується вектором  $\vec{p}_i = p_{1i}, p_{2i}, \dots, p_{ki}$  ознак на області (кластери), що відповідають компактным групам зображень, при цьому вектор ознак характеризується як якісними, так і кількісними ознаками.

Стаття розбита на дві частини: у першій частині розглядається приклад із проведенням процедури кластеризації, як це частково описано в [3], [4], [5], [6], тільки з кількісними ознаками; у другій частині до даних кількісних ознак додаються якісні і показується, яким чином відбувається об'єднання ознак і подальша кластеризація даних.

Приклад вибору оптимальної кластеризації за якісними ознаками.

Вихідна вибірка даних представлена в таблиці 1.

Таблиця 1

Ознаки Об'єкти	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$
$s_1$	5	979	3,7	16	0,9
$s_2$	5	2860	1,8	32	3,4
$s_3$	5	2669	19,1	3	2,0
$s_4$	2	2989	6,4	9	0,9
$s_5$	2	3202	9,4	6	0,7
$s_6$	6	2668	19,4	3	2,2
$s_7$	5	2736	2,4	24	3,6
$s_8$	5	2668	17,4	3	1,9
$s_9$	5	2670	9,8	6	2,8
$s_{10}$	2	1005	3,7	16	1,0
$s_{11}$	2	3953	4,3	14	0,7

**Блок 1.** Нормування змінних. У [1] наведені найбільш розповсюджені способи нормування показників (перехід від вихідних значень  $x$  до нормованого  $z$ ).

© С.С. Бучик, Б.М. Герасимов, 2002

$$z_1 = \frac{x_1 - \mu}{\sigma}$$

(1)

$$z_2 = \frac{x}{x'} \tag{2}$$

$$z_3 = \frac{x}{x'} \tag{3}$$

$$z_4 = \frac{x}{x_{max}} \tag{4}$$

$$z_5 = \frac{(x - \bar{x})}{x_{max} - x_{min}} \tag{5}$$

де  $\bar{x}, \sigma$  – відповідно середнє і середнє квадратичнє відхилення  $x$ ;  $x'$  – деяке еталонне (нормативне) значення  $x$ ;  $x_{max}, x_{min}$  – найбільше і найменше значення  $x$ .

Пронормуємо вихідні дані за формулою (4). Таблиця з нормованими вихідними даними представлена нижче (табл. 2).

Таблиця 2

Ознаки Об'єкти	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$
$s_1$	0,8333	0,2477	0,1907	0,5000	0,2500
$s_2$	0,8333	0,7235	0,0928	1	0,9444
$s_3$	0,8333	0,6752	0,9845	0,0938	0,5556
$s_4$	0,3333	0,7561	0,3299	0,2813	0,2500
$s_5$	0,3333	0,8100	0,4845	0,1875	0,1944
$s_6$	1	0,6749	1	0,0938	0,6111
$s_7$	0,8333	0,6921	0,1237	0,7500	1
$s_8$	0,8333	0,6749	0,8969	0,0938	0,5278
$s_9$	0,8333	0,6754	0,5052	0,1875	0,7778
$s_{10}$	0,3333	0,2542	0,1907	0,5000	0,2778
$s_{11}$	0,3333	1	0,2216	0,4375	0,1944

**Блок 2.** Одержуємо матрицю близькості  $D$ , елементи якої визначаємо як міжточкові відстані, за формулою:

$$l_{ij} = \sqrt{\frac{1}{n} \sum_{k=1}^n (p_{ik} - p_{jk})^2} \tag{6}$$

де  $n$  – кількість всіх ознак (для нашого прикладу  $n = 5$ );

$i = \overline{1, s}$  (для нашого прикладу  $s = 11$ );

$j = \overline{(i + 1), s}$ .

До прикладу з табл. 2:

$$l_{12} = \sqrt{\frac{1}{5} \left[ (0,8333 - 0,8333)^2 + (0,2477 - 0,7235)^2 + (0,1907 - 0,0928)^2 + (0,5 - 1)^2 + (0,25 - 0,9444)^2 \right]} = 0,4401. \tag{7}$$

Розраховуючи аналогічним чином інші міжточкові відстані для нормованих вихідних даних, одержимо таблицю міжточкових відстаней між об'єктами (табл. 3).І

Таблиця 3

S	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
S1	0,4401	0,4629	0,3393	0,3882	0,4818	0,4067	0,4298	0,3628	0,2240	0,4059
S2	0	0,5950	0,5111	0,5716	0,5976	0,1162	0,5734	0,4148	0,4846	0,4944
S3		0	0,4034	0,3626	0,0789	0,5234	0,0411	0,2400	0,5098	0,4870
S4			0	0,0880	0,4617	0,4645	0,3716	0,3390	0,2529	0,1405
S5				0	0,4268	0,5214	0,3341	0,3489	0,3162	0,1831
S6					0	0,5249	0,0952	0,2487	0,5572	0,5381
S7						0	0,5004	0,3199	0,4539	0,4692
S8							0	0,2120	0,4803	0,4563
S9								0	0,4181	0,4095
S10									0	0,3370

Блок 3. Побудова диполів (пар близько розташованих об'єктів) і поділ даних на підвибірки  $A \cap B$  і  $C \cap D$ . Щоб досягти однозначного вибору (регуляризації) несуперечливої кластеризації, останні знаходяться на двох підвибірках даних, на яких потрібний збіг вибору кластерів.

Для вибору кластеризації як основного [4] вихідного критерію використовується критерій несуперечності: цей критерій вимагає, щоб кластеризація, отримана на половині вибірки  $A(C)$ , якнайменше (за всіма показниками) відрізнялася від кластеризації, отриманої на половині вибірки  $B(D)$ . Пояснимо правило утворення диполів, що описане в [5], [7]: за матрицею близькості об'єктів  $D(s_i, s_j)$  будуються диполі, причому побудова диполів продовжується доти, доки всі об'єкти ввійдуть у множини  $A$  і  $B$ , при цьому протилежні вершини (об'єкти) відносяться до різних множин.

Найкоротші диполі наведені нижче:

- 1) 3 0,0411 8                      2) 3 0,0789 6                      3) 4 0,0880 5                      4) 6 0,0952 8
- 5) 2 0,1162 7                      6) 4 0,1405 11                      7) 5 0,1831 11                      8) 8 0,2120 9
- 9) 1 0,2240 10                      10) 3 0,2400 9                      11) 3 0,2400 9                      12) 4 0,2529 10
- .....

У результаті одержимо дев'ять диполів:

Адреса диполя	I	II	III	IV	V	VI	VII	VIII	IX
A:	3	3	4	6	2	4	5	8	1
B:	8	6	5	8	7	11	11	9	10

Блок 4. Обчислюємо таблиці міжточкових відстаней для підвибірок  $A$  (табл. 4) і  $B$  (табл. 5).

Таблиця 4

Підмножина		I	II	III	IV	V	VI	VII	VIII	IX
A		3	3	4	6	2	4	5	8	1
I	3	0	0	0,4034	0,0789	0,5950	0,4034	0,3626	0,0411	0,4629
II	3		0	0,4034	0,0789	0,5950	0,4084	0,3626	0,0411	0,4629
III	4			0	0,4617	0,5111	0	0,0880	0,3716	0,3393
IV	6				0	0,5976	0,4617	0,4268	0,0952	0,4818
V	2					0	0,5111	0,5716	0,5734	0,4401
VI	4						0	0,0880	0,3716	0,3393
VII	5							0	0,3341	0,3882
VIII	8								0	0,4298
IX	1									0

Таблиця 5

Підмножина B		I	II	III	IV	V	VI	VII	VIII	IX
		8	6	5	8	7	11	11	9	0
I	8	0	0,0952	0,3341	0	0,5004	0,4563	0,4563	0,2120	0,4803
II	6		0	0,4268	0,0952	0,5249	0,5381	0,5381	0,2487	0,5572
III	5			0	0,3341	0,5214	0,1831	0,1831	0,3489	0,3162
IV	8				0	0,5004	0,4563	0,4563	0,2120	0,4803
V	7					0	0,4692	0,4692	0,3199	0,4539
VI	11						0	0	0,4095	0,3370
VII	11							0	0,4095	0,3370
VIII	9								0	0,4181
IX	10									0

Блок 5. За таблицями будуюмо два дерева перебору гіпотез про число кластерів (рис. 1).

Блок 6. Здійснюємо перебір кластеризацій за критерієм несуперечності, який розраховуємо за формулою [8]:

$$CY_{AB} = \frac{k - \Delta k}{k} \rightarrow \min, \tag{8}$$

де  $k$  – кількість кластерів, виділена на вибірках  $A$  і  $B$  (різних і тих, що збігаються);

$\Delta k$  – кількість кластерів, що збігаються.

Як видно з рис. 1,  $CY_{AB} = 0$  при  $k_A = k_B = 4$ , що і буде в даному випадку оптимальною кластеризацією. Крайні значення, коли  $k_A = k_B = 1, k_A = k_B = 9$ , є тривіальними і не розглядаються.

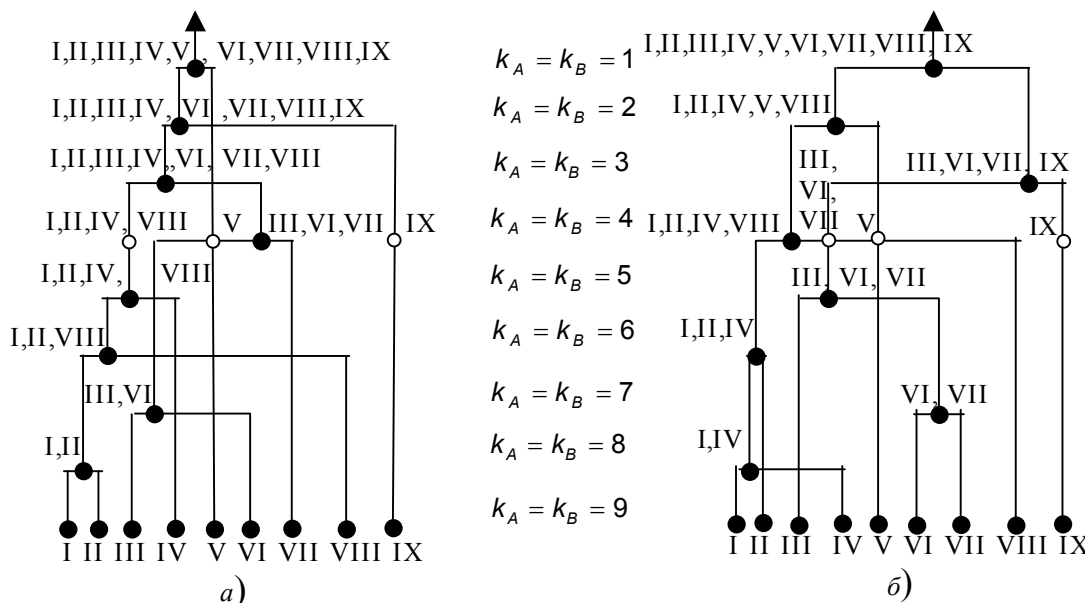


Рис. 1. Ієрархічний перебір гіпотез про число кластерів: а – для підвибірки A; б – для підвибірки B

Запишемо об'єднані адреси об'єктів, при яких  $CY_{AB} = 0$  (табл. 6).

Блок 7. Довизначення (регуляризація). Як зазначено в [5], критерію може бути властива неоднозначність. Він вказує своїм мінімумом оптимальну модель, а також деякі помилкові моделі. Проведемо додаткову перевірку, розділивши множину об'єктів на підмножини C і D, для цього формування підмножин C і D здійснимо за наступним правилом [7]: для диполів, що мають нечіткі номери, об'єкти, що належать A, відносяться до C; B – відповідно до D, а для парних номерів диполів – навпаки. Далі здійснюємо перебір кластеризацій за сумарним критерієм несуперечності, вираз якого має вигляд [7]:

$$CY_{\Sigma} = \frac{1}{2}(CY_{AB} + CY_{CD}) \rightarrow \min. \tag{9}$$

Таким чином, для підмножин C і D одержимо наступні диполі:

Адреса диполя	I	II	III	IV	V	VI	VII	VIII	IX
C:	3	6	4	8	2	11	5	9	1
D:	8	3	5	6	7	4	11	8	10

Далі алгоритм ідентичний як для підвибірок А і В. Наведемо лише дендограми для підвибірок С і D для наочності (рис. 2).

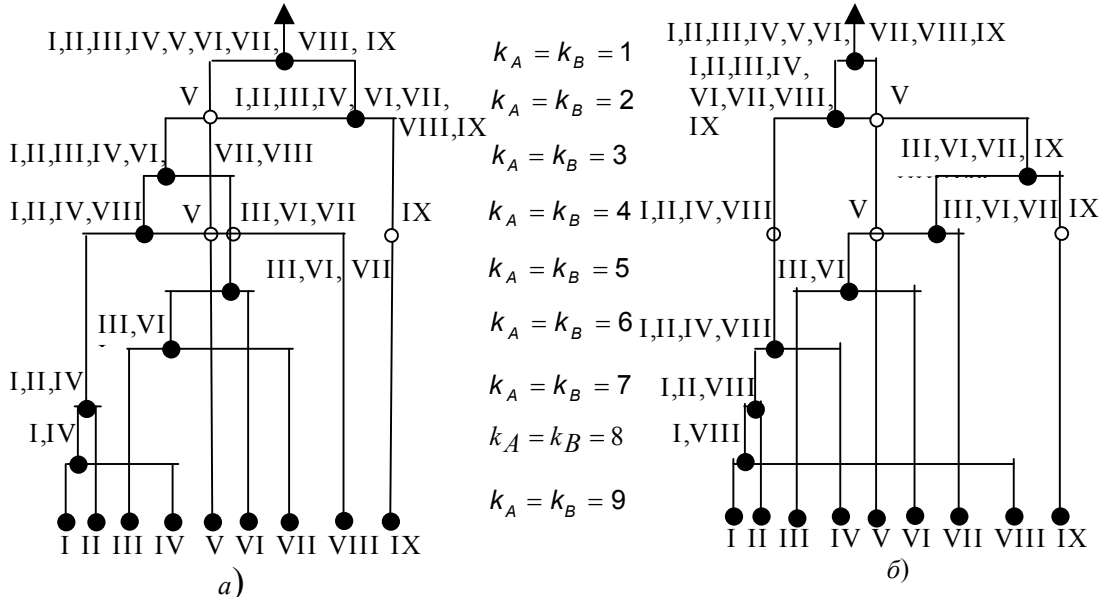


Рис. 2. Ієрархічний перебір гіпотез про число кластерів:  
а – для підвибірки С; б – для підвибірки D

Надалі робимо ті ж операції, що й у блоці 6. Як видно з рис. 2,  $CY_{CD} = 0$  при  $k_C = k_D = 2$  і  $k_C = k_D = 4$ , тобто в даному випадку наявність декількох нулів значень критерію несуперечності. Розраховуючи сумарний критерій  $CY_{\Sigma}$ , видно, що  $CY_{\Sigma} = 0$  при  $k_A = k_B = k_C = k_D = 4$ , що і буде відповідати оптимальній кластеризації (табл. 6).

Таблиця 6

Адреса	Диполі	Об'єкти, об'єднані в кластер	Номер кластера
I II VIII IV	3-8 3-6 8-9 6-8	3, 6, 8, 9	I
V	2-7	2, 7	II
III VI VII	4-5 4-11 5-11	4, 5, 11	III
IX	1-10	1, 10	IV

Приклад вибору оптимальної кластеризації за якісними і кількісними ознаками.

Інформація про якісні ознаки надходить від фахівців-експертів і не є результатом будь-яких об'єктивних спостережень, на відміну від кількісної інформації, що отримана в результаті спостережень за об'єктами. Інформація про можливі значення в якісній формі моделюється лінгвістичними змінними, відповідно до чого якісна ознака  $\rho_{ki}$ , що представляє собою нечітку множину, буде представлена у такий спосіб на кінцевій універсальній множині  $X = \{x_1, x_2, \dots, x_j\}$  [8]:

$$\rho_{ki}^T = \mu^{\rho_{ki}^T}(x_1) / x_1 + \mu^{\rho_{ki}^T}(x_2) / x_2 + \dots + \mu^{\rho_{ki}^T}(x_j) / x_j, \tag{10}$$

де “ $\mu^{p_{ki}}(x_j) / x_j$ ” – пари “функція належності/елемент”, називаються синглтоном; а “+” – позначає сукупність пар; Г – нечіткий терм (наприклад, Н – низький, С – середній, В – високий).

Доповнимо вихідну вибірку даних, представлену в таблиці 1, якісними показниками  $p_6, p_7$  у вигляді терм-множин (табл. 7).

Таблиця 7

Ознаки Об'єкти	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$p_7$
$s_1$	5	979	3,7	16	0,9	н	с
$s_2$	5	2860	1,8	32	3,4	н	в
$s_3$	5	2669	19,1	3	2,0	с	с
$s_4$	2	2989	6,4	9	0,9	с	н
$s_5$	2	3202	9,4	6	0,7	н	н
$s_6$	6	2668	19,4	3	2,2	в	н
$s_7$	5	2736	2,4	24	3,6	н	с
$s_8$	5	2668	17,4	3	1,9	с	с
$s_9$	5	2670	9,8	6	2,8	с	в
$s_{10}$	2	1005	3,7	16	1,0	н	н
$s_{11}$	2	3953	4,3	14	0,7	с	с

Відповідно до (10) нехай якісні ознаки представлені в наступному вигляді:

$$p_6^H = 0,7 / x_1 + 0,2 / x_2 + 0 / x_3 + 0,6 / x_4 + 0,5 / x_5 + 1 / x_6 + 0 / x_7; \tag{11}$$

$$p_6^C = 0,2 / x_1 + 0 / x_2 + 0 / x_3 + 0,6 / x_4 + 0,8 / x_5 + 0,4 / x_6 + 1 / x_7; \tag{12}$$

$$p_6^B = 0,8 / x_1 + 0,7 / x_2 + 0,9 / x_3 + 0,5 / x_4 + 0,6 / x_5 + 0,5 / x_6 + 0,4 / x_7; \tag{13}$$

$$p_7^H = 1 / y_1 + 0,5 / y_2 + 0,5 / y_3 + 0,2 / y_4; \tag{14}$$

$$p_7^C = 0 / y_1 + 0,3 / y_2 + 0,7 / y_3 + 1 / y_4; \tag{15}$$

$$p_7^B = 0,3 / y_1 + 1 / y_2 + 0,5 / y_3 + 0,8 / y_4. \tag{16}$$

**Блок 1.** Нормування кількісних змінних, як це описано в попередньому прикладі (дивитися табл. 2 для кількісних ознак).

**Блок 2.** Одержуємо матрицю близькості D, елементи якої розраховуємо:

– для якісних оцінок знаходимо відносну евклідову відстань  $\varepsilon$  за ступенями приналежності [9], якщо  $A$  і  $B$  – дві нечіткі підмножини, то

$$\varepsilon(A, B) = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mu_A(x_i) - \mu_B(x_i))^2}, \tag{17}$$

де  $n$  – кількість елементів підмножини  $i$ .

$$0 \leq \varepsilon(A, B) \leq 1. \tag{18}$$

Таким чином, можна розрахувати відносні евклідові відстані для якісних показників. До прикладу з табл. 7:

$$\varepsilon(p_{66}^B, p_{66}^C) = \sqrt{\frac{1}{7} \left[ \begin{aligned} &(0,8 - 0,2)^2 + (0,7 - 0)^2 + \\ &+ (0,9 - 0)^2 + (0,5 - 0,6)^2 + \\ &+ (0,6 - 0,8)^2 + (0,5 - 0,4)^2 + \\ &+ (0,4 - 1)^2 \end{aligned} \right]} = 0,5451, \tag{19}$$

$$\varepsilon(p_{47}^H, p_{87}^C) = \sqrt{\frac{1}{4} \left[ (1-0)^2 + (0,5-0,3)^2 + (0,5-0,7)^2 + (0,2-1)^2 \right]} = 0,6558, \tag{20}$$

– узагальнену міжточкову відстань  $\mu(s_i, s_j) = l_{ij}$  знаходимо за наступною формулою:

$$\mu(s_i, s_j) = l_{ij} = \sqrt{\frac{1}{n} \left[ \sum_{k=1}^r (p_{ik} - p_{jk})^2 + \sum_{k=r+1}^n \varepsilon(p_{ik}^T, p_{jk}^T) \right]}, \tag{21}$$

де  $n$  – кількість всіх ознак (для нашого прикладу  $n = 7$ );

$r$  – кількість кількісних ознак (для нашого прикладу  $r = 5$ ).

Розрахувавши за формулою (21) усі міжточкові відстані, одержимо таблицю міжточкових відстаней між об'єктами, яка має наступний вигляд (табл. 8):

Таблиця 8

S	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
S1	0,4431	0,4736	0,4971	0,4487	0,5385	0,3437	0,4507	0,4725	0,3599	0,4347
S2	0	0,6182	0,5768	0,5552	0,5668	0,2601	0,6034	0,4406	0,4926	0,5512
S3		0	0,4582	0,5087	0,3746	0,5166	0,0347	0,3149	0,5921	0,4116
S4			0	0,2770	0,5523	0,5648	0,4385	0,3962	0,3419	0,3283
S5				0	0,5207	0,5365	0,4946	0,4828	0,2672	0,4346
S6					0	0,5666	0,3773	0,3494	0,6024	0,5854
S7						0	0,5001	0,4498	0,4908	0,4780
S8							0	0,3002	0,5741	0,3856
S9								0	0,5206	0,4217
S10									0	0,4960

Далі, як у блоці 3 попереднього прикладу, будемо диполі і розділяємо дані на дві підвибірки  $A \cap B$  і  $C \cap D$ .

Наведемо найкоротші диполі:

- |               |                |                 |                |
|---------------|----------------|-----------------|----------------|
| 1) 3 0,0347 8 | 2) 2 0,2601 7  | 3) 5 0,2672 10  | 4) 4 0,2770 5  |
| 5) 8 0,3002 9 | 6) 3 0,3149 9  | 7) 4 0,3283 11  | 8) 4 0,3419 10 |
| 9) 1 0,3437 7 | 10) 6 0,3494 9 | 11) 1 0,3599 10 | 12) 3 0,3746 6 |
| .....         | .....          | .....           | .....          |

У результаті одержимо десять диполів:

Адреса диполя	I	II	III	IV	V	VI	VII	VIII	IX	X
A:	3	2	5	4	8	3	4	4	1	6
B:	8	7	10	5	9	9	11	10	7	9
C:	3	7	5	5	8	9	4	10	1	9
D:	8	2	10	4	9	3	11	4	7	6

Обчислюємо таблиці міжточкових відстаней для підвбірок A і B, C і D, будемо дерева перебору гіпотез про число кластерів і розраховуємо критерій несуперечності, як це описано в прикладі для кількісних показників.

Для наочності наведемо лише дендограми для підвбірок A і B (рис. 3), C і D (рис. 4).

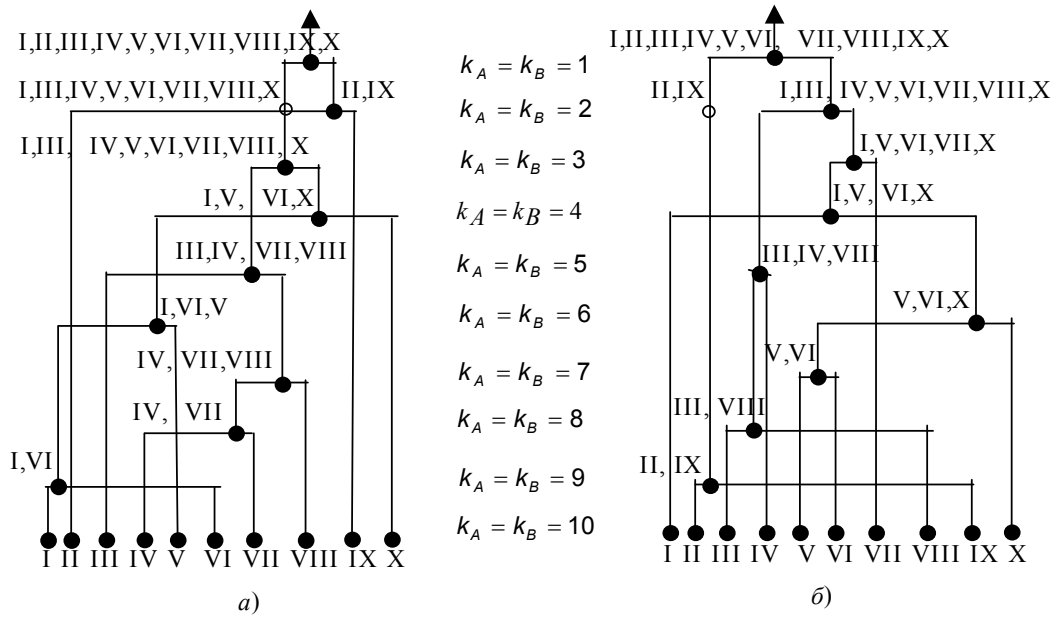


Рис. 3. Ієрархічний перебір гіпотез про число кластерів:  
 а – для підвибірки А; б – для підвибірки В

Як видно з рис. 3,  $CY_{AB} = 0$  при  $k_A = k_B = 2$ , що є в даному випадку оптимальною кластеризацією. Запишемо об'єднані адреси об'єктів, при яких  $CY_{AB} = 0$  (табл. 9).

Таблиця 9

Адреса	Диполі	Об'єкти, об'єднані в кластер	Номер кластера
II IX	2-7 1-7	1, 2, 7	I
I III IV V VI VII VIII X	3-8 5-10 4-5 8-9 3-9 4-11 4-10 6-9	3, 4, 5, 6, 8, 9, 10, 11	II

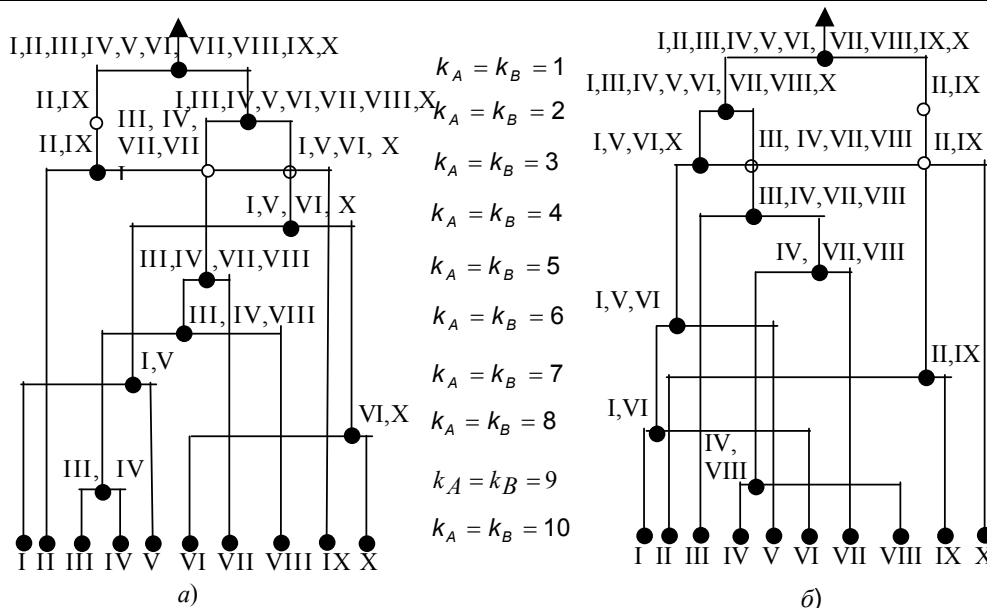




Рис. 4. Ієрархічний перебір гіпотез про число кластерів:  
*a* – для підвибірки *C*; *b* – для підвибірки *D*

Як видно з рис. 4,  $CY_{CD} = 0$  при  $k_C = k_D = 2$  і  $k_C = k_D = 3$ , тобто маємо кілька нулів. Розраховуючи сумарний критерій  $CY_{\Sigma}$ , видно, що  $CY_{\Sigma} = 0$  при  $k_A = k_B = k_C = k_D = 2$ , що і буде відповідати оптимальній кластеризації (табл. 9).

На прикладах показано, яким чином можливо об'єднати інформацію, надану експертами за якісними ознаками, що формалізовані математичним апаратом теорії нечітких множин і кількісними ознаками, отриманими експериментально. Далі на основі застосування алгоритму самоорганізації на основі критерію несуперечності показана можливість кластеризації об'єктів за описаними вище ознаками, при цьому результати кластеризації з включенням якісних ознак можуть уточнювати кластеризацію, отриману лише за кількісними ознаками. Це вказує шлях на побудову програм з питань кластеризації з врахуванням як експериментальних даних, так і думок фахівців-експертів.

#### ЛІТЕРАТУРА:

1. *Мандель И.Д.* Кластерный анализ.– М.: Финансы и статистика, 1988. – 176 с.
2. Нечеткие множества и теория возможностей. Последние достижения: Пер. с англ./ Под ред. Р.Р. Ягера. – М.: Радио и связь, 1986. – 408 с.
3. *Ивахненко А.Г.* Непрерывность и дискретность. – Киев: Наукова думка, 1990. – 224 с.
4. *Ивахненко А.Г.* Объективная кластеризация на основе теории самоорганизации моделей // Автоматика. – 1987. – № 5. – С. 6–15.
5. *Ивахненко А.Г., Ивахненко Н.А.* Непараметрические прогнозирующие модели МГУА. Часть 2. Индикативные системы переборного моделирования, кластеризация распознавания образов // Автоматика. – 1989. – № 2. – С. 3–12.
6. *Ивахненко Н.А., Семина Л.П., Чихрадзе Т.А.* Модифицированный алгоритм объективной компьютерной кластеризации данных//Автоматика. – 1986. – № 2. – С. 10–19.
7. *Герасимов Б.М., Грабовский Г.Г., Рюмишин Н.А.* Нечеткие множества в задачах проектирования, управления и обработки информации. – К.: Техніка, 2002. – 140 с.
8. *Ротштейн А.П.* Интеллектуальные технологии идентификации. – Винница: Універсум-Вінниця, 1999. – 320 с.
9. *Кофман А.* Введение в теорию нечетких множеств. – М.: Радио и связь, 1982. – 432 с.

БУЧИК Сергій Степанович – ад'юнкт Житомирського військового інституту радіоелектроніки ім. С.П. Корольова .

Наукові інтереси:

- розпізнавання нелінійних об'єктів.

ГЕРАСИМОВ Борис Михайлович – доктор технічних наук, професор, провідний науковий співробітник науково-дослідного відділу Житомирського військового інституту радіоелектроніки ім. С.П. Корольова.

Наукові інтереси:

- розробка інтелектуальних систем;
- системи відображення інформації.

Подано 6.06.2002

**Бучик С.С., Герасимов Б.М.** Нечітка кластеризація даних на основі принципу самоорганізації

**Бучик С.С., Герасимов Б.М.** Нечеткая кластеризация данных на основе принципа самоорганизации.

**Buchik S.S., Gerasimov B.M.** Fuzzy clustering of data on the basis of the principle of self-organization.

С.С.Бучик, Б.М.Герасимов. Нечеткая кластеризация данных на основе принципа самоорганизации.

Рассмотрена задача разделения объектов, характеризующихся количественными, полученными экспериментально и качественными, которые предоставлены экспертами и формализованы математическим аппаратом теории нечетких множеств признаками, на группы (кластеры).

S.S.Buchik, B.M.Gerasimov. Fuzzy clustering of data on the basis of the principle of self-organization.

The task of division by groups (clusters) of objects, which are described by quantitative, received experimentally, and by qualitative, being represented by the experts and formalized mathematically by the theory of fuzzy sets attributes, is considered.