

можуть тільки ті ВНЗ, які зможуть інтегруватися й співпрацювати з європейськими закладами вищої освіти в напрямку розробки нових навчальних планів, розробки схем мобільності та інтегрованих програм навчання і досліджень. Органи державного управління освітою, працюючи над сприянням європейській співпраці в забезпеченні якості освіти, закладають основу створенню принципово нової моделі економічної безпеки України. Мобільність студентів і викладачів як одна з цілей Болонського процесу й прагнення її досягнення відкриває можливість для українських спеціалістів стати більш активними суб'єктами міжнародного ринку праці.

Основними джерелами фінансових ресурсів державних ВНЗ можуть бути: надходження за виконання науково-дослідних контактів; надходження від комерційних структур ВНЗ; фонди, отримані від уряду (загальнодержавний бюджет) для фінансування наукових досліджень (при збільшенні частки і ролі грантів); стипендії та виплати студентам, слухачам і аспірантам за навчання; надходження від спонсорів (наглядових рад), трасових фондів, добровільні внески.

Джерела фінансування недержавних прибуткових ВНЗ: плата за навчання; державні субсидії та субсидії з бюджетів місцевих рівнів; окремі внески приватних осіб (які можуть надавати пільги щодо податкових пільг або інших переваг); благодійні внески асоціацій, спілок та інших об'єднань, фірм, підприємств. З одного боку, ВНЗ прагнуть до самостійності, але з іншого — свідомо і несвідомо потребують фінансової та іншої підтримки як з боку органів влади на місцях, так і з боку державних програм. Ця обставина змушує шукати нові форми взаємодії суб'єктів ринку освітніх послуг та інтелектуальних послуг (як замовників, так і споживачів).

Отже, в результаті формування інфраструктури ринку виникає зміна форм власності ВНЗ в економічному їх змісті. Недостатнє фінансування діяльності ВНЗ є джерелом безробіття науково-педагогічних працівників і спонукає ВНЗ шукати додаткові позабюджетні джерела фінансування, тобто надавати в арену приміщення, займатися комерційною і, в тому числі, непрофільною діяльністю, що може призвести до руйнування системи, яка проявляється в редукції цільової функції.

Фінансування фундаментальних і пошукових наукових досліджень, наукових програм, проектів державного значення у вищих навчальних закладах, науково-дослідних установах системи вищої освіти повинні стати пріоритетними напрямками для органів державної влади та місцевого самоврядування. Проте глибока криза, якої зазнавала економіка України останніми десятиліттями, зниження за роки незалежності більш ніж удвічі валового внутрішнього продукту не дають змоги практично реалізувати норми освітнього закону щодо фінансування закладів освіти, актуалізують потребу в глибокому реформуванні національної системи освіти в напрямку кардинального зниження її витратності, піднесення ефективності.

Література: 1. Майминас З. Е. Процесс планирования в экономике: Информационный аспект. — М., 1971. 2. Михайлик І. Державне регулювання економіки / І. Михайлик, Н. Мельник, М. Крупка, З. Залого; [За ред. докт. екон. наук, професора, академіка І. Р. Михайлика. — Львів: Українські технології, 1999. — 640 с. 3. Майборода В. К. Вища педагогічна освіта в Україні: історія, досвід, уроки (1917 — 1985 рр.) / За ред. В. І. Лугового. — К.: Либідь, 1992. — 196 с. 4. Малиновський В. Я. Державне управління: Навчальний посібник. — 2-ге вид., доп. та перероб. — К.: Атіка, 2003. — С. 158. 5. Закон України "Про вищу освіту" // Освіта України. — 2002. 6. Луговий В. І. Педагогічна освіта в Україні: структура, функціонування, тенденції розвитку / За заг. ред. акад. О. Г. Мороза. — К.: МАУП, 1994. — 192 с.

УДК 330.43

**Єгоршин О. О.
Малярець Л. М.**

УНІВЕРСАЛЬНИЙ МЕТОД БАГАТОВИМІРНОГО ШКАЛЮВАННЯ

The universal algorithm of metric and immetric scaling, that has mathematic grounding is woked out in the article.

Мета багатовимірного шкалювання — визначення взаємного розташування об'єктів у деякому теоретичному просторі невеликої розмірності, що адекватно відображає реальність. Таким чином, мета багатовимірного шкалювання фактично співпадає з метою кластерного аналізу — виявлення структури взаємного розміщення об'єктів, виявлення числа та складу груп близьких об'єктів (кластерів), оцінка за деякою шкалою відносних відстаней між об'єктами. Проте поставлене завдання вирішується цими двома багатовимірними статистичними методами зовсім по-різному.

Зауважимо, що методи кластерного аналізу і дотепер не мають ніякого теоретичного обґрунтування — більш ніж піввіковою практикою були відсіянні явно невдалі техніки, що приводили до неправильної класифікації об'єктів; наразі ж відпрацьовані відносно надійні процедури, які витримали ряд перевірок на прикладах аналізу сукупності об'єктів з відомими класифікаціями. Серед ієрархічних процедур кластерного аналізу сьогодні найперспективнішою вважається процедура Уорда.

За вихідними значеннями деякого набору ознак важко безпосередньо скласти образ об'єкта та адекватно його ідентифікувати. В багатовимірному шкалюванні, подібно методу головних компонент, будують деякий скорочений простір узгальнених координат (латентних змінних), у якому проблема ідентифікації вирішується більш просто і надійно в порівнянні з початковим набором ознак, навіть коли ці ознаки метричні. Отже, розрізняють метричне і неметричне багатовимірне шкалювання в залежності від виду ознак — метричних або порядкових. Найбільш обґрунтованим серед метричних методів багатовимірного шкалювання є алгоритм Торґенсона [1; 2].

З обчислювальної точки зору метод Торґенсона надзвичайно схожий з методом головних компонент факторного аналізу, проте між цими методами наявні істотні відмінності. Порівняємо ці методи, оскільки кінцеві результати, що обчислені за ними, отримуються дуже схожими.

У методі головних компонент за матрицею метричних даних X розміру $n \times m$ (n — об'єктів, m — ознак) складають кореляційну матрицю R розміру $m \times m$; знаходять усі її власні числа λ і власні вектори U . На цьому етапі визначають число головних компонент ($p < m$), що достатнє для адекватного подання початкового простору ознак. Далі обчислюються елементи матриці факторних навантажень A розміру $m \times p$ і матриці значень головних компонент Φ розміру $n \times p$. У скороченому просторі головних компонент (латентних факторів) чіткіше проявляються всі основні властивості системи "ознаки — об'єкти", що приховані в таблиці початкових даних. Якщо число головних компонент дорівнює двом ($p = 2$), то розподіл об'єктів можна встановити візуально, зобразивши графік розкиду точок (для кожного об'єкта) в просторі головних компонент. У факторному аналізі (вже для інших цілей) додатково застосовується перетворення обертання, що переводить факторну матрицю A в нову матрицю факторних навантажень B , а матрицю значень головних компонент Φ — у нову матрицю значень головних факторів F (розміри матриць не змінюються). На взаємне розміщення об'єктів у факторному просторі ця операція обертання не впливає.

У методі Торнгенсона (основному методі багатовимірного шкалювання) за матрицею метричних даних X розміру $n \times m$ складають матрицю квадратів евклідових відстаней D^2 розміру $n \times n$. Елементи цієї матриці обчислюються за відомою формулою:

$$d_{ij}^2 = \sum_{k=1}^m (x_{ik} - x_{jk})^2. \quad (1)$$

Зауваження. Попередньо значення всіх ознак повинні бути нормовані.

Для матриці D^2 обчислюють середні по рядках $d_{i\bullet}^2$, по стовпцях $d_{\bullet j}^2$ і загальне середнє $d_{\bullet\bullet}^2$; далі за допомогою подвійного центрування переходять до матриці D^* , елементи якої визначають за формулою:

$$d_{ij}^* = -\frac{1}{2}(d_{ij}^2 - d_{i\bullet}^2 - d_{\bullet j}^2 + d_{\bullet\bullet}^2). \quad (2)$$

Подальші викладки майже повністю повторюють етапи методу головних компонент. Для матриці D^* знаходять власні числа μ і власні вектори V . Матриця D^* має істотно більші розміри, ніж кореляційна матриця R , але число ненульових власних чисел у цих матриц однакове і не перевищує числа ознак m . Головні "компоненти" для матриці D^* у багатовимірному шкалюванні називаються "стимулами", їх число виявляється тим самим, що число головних компонент з матрицею R (для визначення числа головних компонент використовують деякі евристичні правила, тому тут можливі невеликі відхилення).

Матриця стимулів S складається аналогічно матриці факторних навантажень A , але за розмірами $n \times p$ і змістом вона еквівалентна матриці головних компонент Φ . У скороченому просторі стимулів (латентних змінних) чіткіше проявляються всі властивості системи "ознаки – об'єкти", що приховані в таблиці початкових даних. Якщо число стимулів дорівнює двом ($p = 2$), то розподіл об'єктів можна встановити візуально, зобразивши графік розкиду точок (для кожного об'єкта) в просторі стимулів. Як і у факторному аналізі, допустимі перетворення матриці стимулів методом обертання, але для задачі розпізнавання образів ця операція вже нічого не дає.

Можна довести, що при виборі як норми звичайної стандартизації змінних матриця D^* (що одержана після подвійного центрування матриці квадратів евклідових відстаней) буде пропорційною кореляційній матриці RR між об'єктами (розмір цієї матриці $n \times n$). Нагадаємо, що в методі Торнгенсона в якості норми можливо вибирати будь-які інші міри, наприклад, середні значення ознак. Центрування (оскільки при стандартизації дані центруються) не змінюють евклідових відстаней між об'єктами.

Отже, виникає ідея застосовувати розклад кореляційної матриці RR між об'єктами на суму внесків:

$$RR = \mu_1 V_1 V_1' + \mu_2 V_2 V_2' + \dots + \mu_m V_m V_m', \quad (3)$$

де μ – власні числа;

V – власні вектори кореляційної матриці RR .

Неважко довести, що всі власні числа μ кореляційної матриці – дійсні і невід'ємні, а власні вектори V – взаємно ортогональні (тобто $V_i' V_j = 0$, $V_i' V_i = 1$). У цьому розкладі внесок кожного члена буде пропорційний відповідному власному числу (оскільки норма кожного власного вектора дорівнює одиниці). Декілька перших доданків із найбільшими власними числами відтворюють достатньо точно всю матрицю RR , а доданки з малими власними числами можуть бути відкинуті. Згідно з найпоширенішою рекомендацією, сума залишених у моделі власних чисел повинна бути близькою до 70% від їх загальної суми.

Допускається користування й іншими евристичними рекомендаціями, що прийняті у факторному аналізі. Число залишених у моделі доданків позначимо через p . Це число значно менше порядку матриці n , більш того – менше числа ознак $p < m$.

Таким чином, матрицю RR можна навести у вигляді скалярного добутку:

$$RR = SS', \quad (4)$$

де S – матриця розміру $n \times p$, її елементи обчислюються за формулою:

$$s_{ij} = v_{ij} \sqrt{\mu_j}. \quad (5)$$

Цікавим також є порівняння одержаних результатів з методом головних компонент.

По-перше, кореляційні матриці R і RR (між ознаками та між об'єктами) мають однаковий ранг, тобто число ненульових власних чисел у них одне й те ж. Більш того, можна довести, що ненульові власні числа для матриць $Z'Z$ і ZZ' (що пропорційні кореляційним матрицям R і RR) однакові. По-друге, можна довести, що власні вектори V матриці RR співпадають з головними компонентами Φ для матриці R , а пропорційні їм вектори S співпадають зі "стимулами" за методом Торнгенсона, коли прийнята попередня стандартизація даних. Отже, матриця S – це матриця координат кожного об'єкта в скороченому просторі латентних змінних.

У багатовимірному шкалюванні розшарування об'єктів у скороченому просторі "стимулів" досліджується візуально, для чого штучно приймається $p = 2$ (тобто розглядається двовимірний простір перших двох "стимулів"). Це не дуже добре, оскільки двох координат (компонент, "стимулів") може виявитися не досить, щоб адекватно відобразити реальність. Тому пропонується виділяти стільки координат, скільки треба, а недосконалий візуальний аналіз замінити кластерним аналізом за методом Уорда. У скорочених просторах більш надійно працюють будь-які статистичні методи, зокрема, кластерний аналіз. Кластерний аналіз представляє результати у зручній графічній формі, крім того, він супроводжує висновки наданням усієї необхідної інформації про склад і характеристики кожного кластера.

Якщо вихідні дані є порядковими, класичний метод Торнгенсона не може бути застосований, оскільки для початку аналізу за цим методом потрібно мати матрицю евклідових відстаней між об'єктами, а відстані між рангами не мають сенсу (ранги визначають лише порядок розташування даних). З тієї ж причини не може бути застосований кластерний аналіз безпосередньо на вихідних даних (порядкових).

За порядковими даними можна обчислювати коефіцієнти кореляції, причому якщо дані ранжировані, то ці коефіцієнти є коефіцієнтами рангової кореляції Спірмена. Метод головних компонент з кореляційною матрицею між ознаками в даному випадку нічого не дає, оскільки для обчислення координат головних компонент (за якими виявляється взаємне розташування об'єктів у скорочених просторах) потрібно мати метричну матрицю ознак (а не порядкову).

Але запропонований вище метод з кореляційною матрицею між об'єктами без будь-яких труднощів узагальнюється на порядкові дані. Незважаючи на те, що вихідні (порядкові) дані не мають метрики для кількісних порівнянь, наприкінці будуть одержані координати об'єктів у скороченому метричному просторі.

Слід розглянути приклад, де будуть продемонстровані всі вищенаведені теоретичні твердження.

У табл. 1 наведена стандартизована матриця 10 ознак, що кількісно характеризують виробничо-господарську діяльність 7 підприємств за 3 роки (розмір цієї матриці 21×10).

Таблиця 1

Закінчення табл. 3

Стандартизовані значення ознак (матриця Z)

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	0,633	-0,352	0,677	0,527	-0,872	-1,104	1,501	1,226	1,821	1,338
2	-0,108	-0,152	-0,199	0,178	-0,663	-0,967	-0,075	-0,040	-1,655	-0,418
3	-0,233	-0,048	-0,342	0,174	-0,660	-0,973	-0,075	-1,729	0,083	-1,296
4	1,856	-1,717	1,776	-0,303	-0,245	-0,096	1,501	1,226	1,821	1,338
5	1,127	-1,120	0,975	-0,643	0,348	0,202	-0,075	-0,040	-1,655	-0,418
6	1,101	-1,099	0,947	-0,648	0,359	0,184	-2,439	1,226	0,083	-1,296
7	-0,979	1,742	-0,666	-0,823	0,808	-1,368	-0,075	-0,040	1,821	1,338
8	-0,731	0,401	-0,545	-1,061	1,987	-1,325	-0,075	-0,040	0,083	-0,418
9	-1,192	0,779	-1,078	-1,061	1,993	-1,335	-0,075	1,226	0,083	-1,296
10	-0,590	1,378	-0,910	1,566	-1,198	0,114	1,501	1,226	0,083	-0,418
11	-0,585	0,282	0,004	2,228	-1,006	0,548	-0,075	-0,040	0,083	-0,418
12	-0,579	0,276	0,012	2,235	-1,307	0,551	-0,075	-0,040	0,083	-0,418
13	0,411	-0,170	0,555	-0,359	-0,174	-0,469	-0,075	1,226	0,083	1,338
14	-0,185	-0,046	-0,292	-0,884	1,069	-0,407	-0,075	-0,040	0,083	-0,418
15	-0,194	-0,039	-0,298	-0,885	1,071	-0,450	-2,439	-1,729	0,083	-0,418
16	-0,812	1,651	-1,032	-0,624	0,276	-0,087	-0,075	-0,040	0,083	1,338
17	-1,474	1,009	-1,730	-0,749	0,617	0,811	-0,075	-0,040	-1,655	1,338
18	-1,477	1,012	-1,734	-0,750	0,619	0,805	-0,075	-0,040	0,083	-0,418
19	1,523	-1,353	1,532	0,548	-0,882	1,349	1,501	-0,040	0,083	-0,418
20	1,248	-1,219	1,179	0,667	-0,921	2,008	-0,075	-1,729	-1,655	-1,296
21	1,240	-1,213	1,171	0,666	-0,920	2,009	-0,075	-1,729	0,083	1,338

У табл. 2 обчислені власні вектори матриці $Z'Z$ (що пропорційна кореляційній матриці R між ознаками), в останніх рядках таблиці наведені відповідні власні числа λ_i , їх величини у відсотках до загальної суми і накопичені суми власних чисел у відсотках.

Таблиця 2

Матриця власних векторів U для матриці $Z'Z$

№	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
1	0,460	0,038	0,347	-0,007	0,035	0,024	-0,068	-0,313	0,334	0,671
2	-0,440	0,106	-0,362	-0,071	-0,064	-0,073	0,020	-0,270	0,759	-0,014
3	0,467	0,071	0,311	0,049	-0,090	-0,021	-0,091	0,228	0,476	-0,618
4	0,293	0,086	-0,549	0,262	-0,254	-0,162	-0,069	0,577	0,096	0,315
5	-0,389	-0,114	0,434	-0,058	0,152	0,124	0,323	0,623	0,234	0,242
6	0,330	-0,248	-0,250	-0,370	0,272	-0,285	0,684	-0,052	0,030	-0,061
7	0,149	0,469	-0,238	0,028	0,422	0,705	0,150	0,032	0,020	-0,031
8	-0,087	0,465	0,126	0,450	0,479	-0,566	0,071	-0,033	-0,041	-0,019
9	-0,033	0,524	0,164	-0,013	-0,642	-0,003	0,510	-0,104	-0,120	0,004
10	-0,008	0,436	0,013	-0,762	0,078	-0,231	-0,349	0,202	-0,065	0,062
λ	75,99	44,06	36,54	19,13	14,52	9,62	6,98	1,78	1,20	0,81
%	36,2%	21,0%	17,4%	9,1%	6,9%	4,6%	3,3%	0,8%	0,6%	0,1%
Σ	36,2%	57,2%	74,6%	83,7%	90,6%	95,2%	98,5%	99,3%	99,9%	100,0%

Як видно з табл. 2, перші три власних числа складають близько 75% від загальної суми, тобто досить виділити три головних компоненти. Їх значення $F = ZU$ обчислені в табл. 3 для всіх об'єктів (7 підприємств за 3 роки); нормовані значення компонент позначені через Φ .

Таблиця 3

Значення головних компонент

№	F_1	F_2	F_3	Φ_1	Φ_2	Φ_3
1	2	3	4	5	6	7
1	0,937	3,264	-0,276	0,108	0,492	0,046
2	-0,035	-0,808	0,451	-0,004	-0,122	-0,075
3	-0,111	-1,067	0,511	-0,013	-0,161	-0,085
4	2,459	2,851	-2,012	0,282	0,430	0,333
5	1,259	-1,256	-1,290	0,144	-0,189	0,213
6	0,701	-1,245	-2,273	0,080	-0,188	0,376
7	-2,612	1,760	-0,295	-0,300	0,265	0,049
8	-2,295	-0,205	-1,229	-0,263	-0,031	0,203
9	-3,032	-0,013	-0,920	-0,348	-0,002	0,152
10	-0,223	1,437	2,590	-0,026	0,216	-0,428

1	2	3	4	5	6	7
11	0,942	0,019	2,210	0,108	0,003	-0,366
12	0,955	0,019	2,208	0,110	0,003	-0,365
13	0,199	1,303	-0,818	0,023	0,196	0,135
14	-1,017	-0,322	-0,934	-0,117	-0,048	0,154
15	-1,247	-2,205	-1,290	-0,143	-0,332	0,213
16	-1,922	0,580	0,672	-0,220	0,087	-0,111
17	-2,086	-0,745	1,178	-0,239	-0,112	-0,195
18	-2,136	-0,599	0,917	-0,245	-0,090	-0,152
19	3,187	0,383	-0,118	0,366	0,058	0,019
20	3,082	-2,605	0,514	0,354	-0,392	-0,085
21	2,993	-0,547	0,203	0,343	-0,082	-0,034
СумКв	75,99	44,06	36,54	1	1	1

Для порівняння в табл. 4 обчислені значення перших 10 власних векторів V , що відповідають ненульовим власним числам матриці ZZ' .

Таблиця 4

Власні вектори V матриці ZZ'

№	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
1	0,108	0,492	-0,046	-0,046	-0,112	-0,064	0,155	0,033	0,224	-0,122
2	-0,004	-0,122	0,046	-0,175	0,156	-0,080	0,597	0,107	0,172	0,053
3	-0,013	-0,161	0,085	-0,146	-0,367	-0,450	0,184	0,351	0,274	0,081
4	0,282	0,430	-0,333	0,064	0,049	-0,074	-0,125	-0,038	0,232	-0,098
5	0,144	-0,189	-0,213	-0,040	0,334	-0,080	0,229	-0,070	-0,219	-0,022
6	0,080	-0,188	-0,376	-0,304	-0,080	0,623	-0,121	0,285	0,013	-0,033
7	-0,300	0,265	-0,049	0,222	-0,326	-0,046	0,004	0,102	-0,404	0,368
8	-0,263	-0,031	-0,203	-0,078	0,019	-0,267	-0,034	-0,446	-0,130	-0,111
9	-0,348	-0,002	-0,152	-0,351	0,161	-0,093	-0,215	-0,225	-0,024	0,035
10	-0,026	0,216	0,429	-0,277	0,147	0,023	-0,089	0,246	-0,260	-0,585
11	0,108	0,003	0,366	-0,169	-0,207	0,209	-0,007	-0,344	0,034	0,164
12	0,110	0,003	0,365	-0,169	-0,208	0,210	-0,007	-0,347	0,032	0,164
13	0,023	0,196	-0,135	0,079	0,136	0,282	0,296	0,021	-0,031	0,307
14	-0,117	-0,048	-0,154	-0,033	0,042	-0,149	-0,128	-0,064	0,060	-0,095
15	-0,143	-0,332	-0,213	0,153	-0,435	0,076	0,062	-0,052	0,039	-0,348
16	-0,220	0,087	0,111	0,309	0,036	0,096	0,070	0,261	-0,362	-0,076
17	-0,239	-0,112	0,195	0,388	0,436	0,142	0,092	-0,110	0,309	-0,026
18	-0,245	-0,090	0,152	0,087	0,107	0,012	-0,475	0,291	0,396	0,213
19	0,366	0,058	-0,019	-0,046	0,164	-0,225	-0,281	0,083	-0,132	0,241
20	0,354	-0,393	0,085	0,034	0,093	-0,170	-0,109	0,089	-0,285	0,144
21	0,343	-0,082	0,034	0,498	-0,146	0,027	-0,098	-0,173	0,061	-0,254
λ	75,99	44,06	36,54	19,13	14,52	9,62	6,98	1,78	1,20	0,81
%	36,2%	21,0%	17,4%	9,1%	6,9%	4,6%	3,3%	0,8%	0,6%	0,1%
Σ	36,2%	57,2%	74,6%	83,7%	90,6%	95,2%	98,5%	99,3%	99,9%	100%

Порівнюючи табл. 4 з табл. 3, помічаємо, що власні вектори V_i матриці ZZ' (які відповідають її ненульовим власним числам) дійсно співпадають з нормованими компонентами Φ_i матриці $Z'Z$ (в табл. 3 наведені значення лише перших 3-х головних компонент). У підсумкових рядках табл. 4 і табл. 3 подані власні числа матриць ZZ' і $Z'Z$, звідки видно, що ненульові власні числа обох цих матриць дійсно співпадають.

В обох методах виділяються 3 латентних ознаки – 3 компоненти (в методі головних компонент), чи 3 стимули, координати яких обчислені в табл. 5 за формулою:

$$S_i = V_i \sqrt{\lambda_i} \quad (6)$$

На відміну від головних компонент, стимули не нормовані – їх суми квадратів дорівнюють не одиниці, а власним числам λ_i . Як видно з табл. 5 і з табл. 3, стимули співпадають з ненормованими компонентами: $S = F$.

Таблиця 5

Матриця координат стимулів

№	S_1	S_2	S_3
1	0,937	3,264	-0,276
2	-0,035	-0,808	0,451
3	-0,111	-1,067	0,511
4	2,459	2,851	-2,012
5	1,259	-1,256	-1,290
6	0,701	-1,245	-2,273
7	-2,612	1,760	-0,295
8	-2,295	-0,205	-1,229
9	-3,032	-0,013	-0,920
10	-0,223	1,437	2,590
11	0,942	0,019	2,210
12	0,955	0,019	2,208
13	0,199	1,303	-0,818
14	-1,017	-0,322	-0,934
15	-1,247	-2,205	-1,290
16	-1,922	0,580	0,672
17	-2,086	-0,745	1,178
18	-2,136	-0,599	0,917
19	3,187	0,383	-0,118
20	3,082	-2,605	0,514
21	2,993	-0,547	0,203
СумКв	75,993	44,058	36,538

Переходимо до заключного (візуального) етапу багатовимірного шкалювання. Як було вже сказано вище, на цьому етапі замість візуального аналізу будемо користуватися кластерним аналізом за методом Уорда (найкращим методом кластерного аналізу).

У такому випадку є деякі розбіжності між методом головних компонент і методами багатовимірного шкалювання. В методі головних компонент розшарування об'єктів вивчається в скороченому просторі *нормованих* компонент Φ , а в багатовимірному шкалюванні – у скороченому просторі *ненормованих* стимулів S . У науковій літературі описані випадки, коли при збільшенні відстаней за однією координатою (змінною масштабу) деякі об'єкти можуть виявитися близькими до інших угруповань. Будемо мати це на увазі.

За трьома головними компонентами Φ (табл. 3) методом Уорда побудована дендрограма і виділено 7 основних кластерів (рис. 1). Унизу під дендрограмами наведена таблиця, в якій показано, до якого кластера віднесено кожен об'єкт. Нагадуємо, що в табл. 1 фактично наведені дані за 7-ма основними об'єктами за послідовні 3 роки, тому очікувалося об'єднання трьох послідовних об'єктів (рядків матриці даних) у споріднені групи (кластери), що є реальними щорічними станами основних об'єктів.

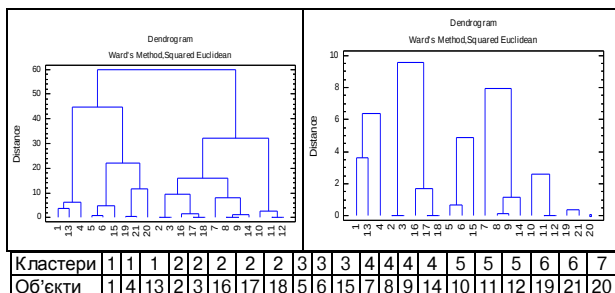


Рис. 1. Кластерний аналіз на головних компонентах (3 компоненти)

За дендрограмами рис. 1 виявилися близькими об'єкти (2, 3); (5, 6); (7, 8, 9); (10, 11, 12); (16, 17, 18); (19, 20, 21). Решта об'єктів 1, 4, 13, 14, 15 (23,8%) віднесені до інших кластерів. Це гарний результат. До речі, чисто візуальний аналіз у просторі перших двох компонент є дещо суб'єктивним і не виявляє таку гарну узгодженість висновків.

За методом багатовимірного шкалювання були виділені 3 стимули, що також як і в методі головних компонент пояснюють 74,6% мінливості початкових даних, і далі за цими стимулами побудовані дендрограми, які нічим не відрізняються від дендрограм, що побудовані на головних компонентах. Ефекту різних масштабів не помічено.

Переходимо до порівняльного аналізу неметричних даних.

Таблицю вихідних даних (табл. 1) ранжируємо за кожним рядком, тобто збережемо в новій таблиці (табл. 6) для кожного об'єкта лише порядок ознак, а не їх кількісні значення. Чи можливо за цими даними уявити, які об'єкти є близькими, а які – ні, і чи є в даному випадку якість угруповання об'єктів на споріднені кластери?

За даними табл. 6 обчислена матриця (розміру 21×21) коефіцієнтів рангової кореляції Спірмена і для цієї матриці знайдені власні вектори та власні числа. В табл. 7 наведені значення перших 10 власних векторів і відповідних їм власних чисел (останні три рядки табл. 7).

Таблиця 6

Таблиця ранжированих даних за рядками

Ранги	X_{16}	X_{17}	X_{18}	X_{34}	X_{35}	X_{38}	C	D	E	F
1	5	3	6	4	2	1	9	7	10	8
2	7	6	5	10	3	2	8	9	1	4
3	6	8	5	10	4	3	7	1	9	2
4	10	1	8	2	3	4	7	5	9	6
5	10	2	9	3	8	7	5	6	1	4
6	9	3	8	4	7	6	1	10	5	2
7	2	9	4	3	7	1	5	6	10	8
8	3	9	4	2	10	1	6	7	8	5
9	3	8	4	5	10	1	6	9	7	2
10	3	8	2	10	1	6	9	7	5	4
11	2	8	6	10	1	9	4	5	7	3
12	2	8	6	10	1	9	4	5	7	3
13	7	4	8	2	3	1	5	9	6	10
14	5	7	4	1	10	3	6	8	9	2
15	7	8	6	3	10	4	1	2	9	5
16	2	10	1	3	8	4	5	6	7	9
17	3	9	1	4	7	8	5	6	2	10
18	2	10	1	3	8	9	5	6	7	4
19	9	1	10	6	2	7	8	4	5	3
20	9	4	8	7	5	10	6	1	2	3
21	8	2	7	6	3	10	4	1	5	9

Таблиця 7

Перші 10 власних векторів U матриці рангових кореляцій

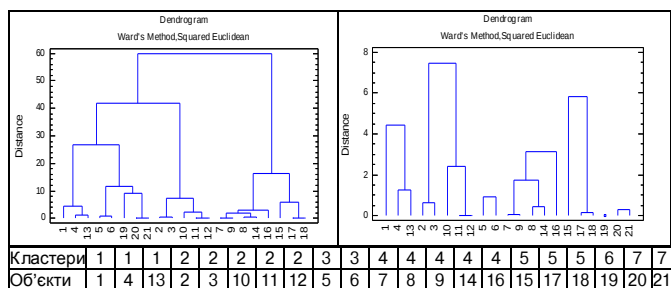
U	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
1	0,044	-0,147	0,495	-0,187	-0,128	-0,045	-0,219	-0,057	-0,148	0,028
2	-0,078	0,105	0,282	0,171	0,485	0,221	0,264	0,289	0,150	0,144
3	0,002	0,207	0,210	0,280	-0,375	0,406	0,172	0,278	0,054	-0,017
4	-0,151	-0,313	0,227	-0,064	-0,247	-0,051	-0,287	0,383	0,107	-0,086
5	-0,211	-0,277	-0,242	0,134	0,215	0,066	-0,055	-0,146	0,253	-0,054
6	-0,090	-0,244	-0,131	0,330	0,148	-0,517	0,170	0,394	-0,166	0,159
7	0,335	-0,056	0,150	-0,121	-0,194	-0,037	0,157	-0,154	0,155	-0,141
8	0,349	-0,133	0,026	0,122	-0,024	0,109	-0,027	-0,228	0,247	0,797
9	0,292	-0,057	0,077	0,372	0,139	0,038	0,010	-0,226	-0,205	-0,302
10	0,018	0,377	0,261	0,028	0,180	0,045	-0,225	0,327	0,051	0,082
11	-0,061	0,391	0,058	0,104	-0,158	-0,399	0,082	-0,128	0,195	0,019
12	-0,061	0,391	0,058	0,104	-0,158	-0,399	0,082	-0,128	0,195	0,019
13	0,017	-0,299	0,283	-0,273	0,115	-0,244	0,266	0,057	0,505	-0,086
14	0,268	-0,200	-0,023	0,302	-0,055	-0,058	-0,368	0,129	0,060	0,002
15	0,153	-0,162	-0,249	0,128	-0,421	0,078	0,389	0,249	0,151	-0,092
16	0,344	0,049	-0,076	-0,247	-0,001	0,058	0,013	0,157	0,076	-0,098
17	0,168	0,141	-0,270	-0,385	0,262	0,069	-0,038	0,173	0,181	-0,088
18	0,256	0,171	-0,270	0,024	-0,045	-0,158	-0,451	0,239	0,190	-0,006
19	-0,335	-0,087	0,141	0,121	-0,092	0,024	-0,274	-0,190	0,277	-0,043
20	-0,307	0,063	-0,248	0,114	-0,074	0,271	-0,098	-0,045	0,376	-0,050
21	-0,275	0,005	-0,156	-0,342	-0,244	0,000	0,042	0,119	-0,266	0,388
λ	6,962	4,668	3,002	2,253	2,056	1,002	0,697	0,211	0,148	0,000
%	33,15%	22,39%	14,30%	10,73%	9,79%	4,77%	3,32%	1,01%	0,71%	0,00%
Cum %	33,15%	55,38%	69,68%	80,41%	90,20%	94,97%	98,29%	99,29%	100,00%	100,00%

$$S_{ij} = u_{ij} \sqrt{\lambda_j}. \quad (7)$$

Координати стимулів

Стимули	S_1	S_2	S_3
1	0,116	-0,318	0,857
2	-0,207	0,226	0,488
3	0,006	0,446	0,365
4	-0,399	-0,677	0,393
5	-0,557	-0,598	-0,420
6	-0,237	-0,526	-0,228
7	0,883	-0,122	0,260
8	0,921	-0,287	0,044
9	0,772	-0,124	0,134
10	0,048	0,815	0,453
11	-0,161	0,845	0,101
12	-0,161	0,845	0,101
13	0,044	-0,645	0,490
14	0,706	-0,431	-0,041
15	0,403	-0,350	-0,431
16	0,908	0,107	-0,132
17	0,443	0,304	-0,469
18	0,675	0,369	-0,468
19	-0,885	-0,188	0,243
20	-0,809	0,137	-0,430
21	-0,726	0,012	-0,271
Сумма	6.962	4.668	3.002

Подальший візуальний аналіз (як і для попередніх прикладів) заміняємо кластерним аналізом у скороченому просторі головних компонентів. Слід нагадати, що безпосередній кластерний аналіз на вихідних порядкових даних неможливий, оскільки за порядковими даними немає можливості обчислити евклідові відстані між об'єктами. На рис. 2 наведені побудовані у скороченому метричному просторі дендрограми і таблиця складу виділених 7-ми кластерів.



Як видно з рис. 2, аналіз правильно визначив розташування більше 70% об'єктів за спорідненими групами: (2, 3); (5, 6); (7, 8, 9); (10, 11, 12); (17, 18); (19, 20, 21). Лише шість об'єктів 1, 4, 13, 14, 15, 16 (28,6%) були віднесені до інших кластерів. Це дуже добрий результат – метричний метод (рис. 1), коли була відома вся кількісна інформація, а не лише порядкова, визначив практично ті самі кластери і розпізнав лише на один об'єкт більше, ніж запропонований авторами неметричний метод багатовимірного шкалювання.

Dendrogram
Ward's Method Squared Euclidean

Distance

Dendrogram
Ward's Method Squared Euclidean

Distance

Кластери	1	1	1	2	2	3	3	4	4	4	5	5	5	6	6	7	7	7			
Об'єкти	1	9	17	2	3	18	4	11	16	5	7	19	6	8	21	10	15	12	13	14	20

Отже, вище запропонований відносно простий універсальний алгоритм як метричного, так і неметричного шкалювання, який має математичне обґрунтування.

Література: 1. Дэйвисон М. Многомерное шкалирование: Методы наглядного представления данных / Пер. с англ. В. С. Каменского. – М.: Финансы и статистика, 1988. – 256 с. 2. Терехина А. Ю. Анализ данных методами многомерного шкалирования. – М.: Наука. 1986. – 168 с.

Стаття надійшла до редакції
15.11.2007 р.

УДК 378.147

Котов А. Н.

РАЗВИТИЕ КРИТИЧЕСКОГО МЫШЛЕНИЯ КАК ФАКТОР СТАНОВЛЕНИЯ КОМПЕТЕНТНОГО СПЕЦИАЛИСТА

In the article the competence approach in higher education system is considered.

Ускорение развития новой техники и технологии, стремительное увеличение объема новых знаний, усложнение общественных и экономических отношений приводит к повышению требований по подготовке квалифицированных спе-