Журнал обчислювальної та прикладної математики | 2016, №3(123) | Journal of Computational & Applied Mathematics

УДК 512.64:004.93
MSC 15Bxx, 68T10

# GROUPING INFORMATION PROBLEM FOR REPRESENTATIVES OF MATRICES AND CONCEPT CORTEGE ITS DECISION

Volodymyr Donchenko, Taras Zinko
Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, E-mail: voldon@unicyb.kiev.ua.

# ЗАДАЧІ ГРУПУВАННЯ ІНФОРМАЦІЇ ДЛЯ МАТРИЧНИХ ПРЕДСТАВНИКІВ ТА КОНЦЕПЦІЯ КОРТЕЖНОСТІ У ЇЇ РОЗВ'ЯЗАННІ

В. С. Донченко, Т. П. Зінько
Факультет комп'ютерних наук та кібернетики, Київський національний університет імені Тараса Шевченка, Київ, Україна, E-mail: voldon@unicyb.kiev.ua.

Abstract. We consider the problem of grouping information, which include the problem of restoring the functions represented their values and the problem of classification, clustering and pattern recognition. Emphasis is placed on the problem of determining eligibility class clusters in problems of classification based on the «immersion» in the natural algebraic structure of Euclidean space. The procedure SVD-filtration matrix considered.
Keywords: pseudoinverse, matrix cortege, grouping information problem, classification, clustering, pattern recognition.

Резюме. Розглядаються задачі групування інформації, які включають в себе задачу відновлення функцій, представлених своїми значеннями та задачі класифікації, кластеризації, та розпізнавання образів. Основна увага приділяється проблемі визначення критеріїв відповідності класам-кластерам в задачах класифікації на основі «занурення» в природні алгебраїчні структури евклідового простору. Розглядається процедура SVD-фільтрації матриць.
Ключові слова: псевдообернення, матричний кортеж, задачі групування інформації, класифікація, кластеризація, розпізнавання образів.

## Вступ

Розглядаються задачі групування інформації (*GIP — Grouping Information Function*), які включають в себе задачу відновлення функцій, представлених своїми значеннями (*function fitting*) та задачі класифікації, кластеризації та розпізнавання образів. Евклідові простори $R^n$, як і загалом в математичному моделюванні, є найбільш вживаною фундаментальною математичною структурою для цього класу задач. Це, з одного боку, визначається природністю використання набору числових характеристик для дослідження об'єктів групування. З іншого — багатством математичних засобів опису структурних зв'язків між числовими векторами та засобів опису базових структур в рамках таких просторів, включаючи матричну алгебру, сингулярне (SVD) подання матриць та псевдообернення (ПдО) за Муром-Пенроузом для матриць. Зокрема, згадані засоби надають можливості «породження» послідовностями числових векторів базових «множинних» структур евклідових просторів числових векторів: підпросторів, зміщених підпросторів (в подальшому — гіперплощин), еліпсів та еліпсоїдальних циліндрів з оптимальними щодо набору властивостями. Конструктивність означає побудову асоційованих з «множинними» структурами матриць, які забезпечують обчислення слушних «відстаней відповідності» «множинним» структурам. Під «*відстанями відповідності*» будемо розуміти чи прямо відстані від підпростору чи гіперплощини, чи слушним чином побудовану «еліпсоїдальну відстань». Згадана конструктивна «асоціативність» дозволяє вводити та досліджувати відповідність об'єктів класам, асоційованим з навчальною вибіркою, через її «занурення» — асоціацію у слушну множинну структуру $R^n$[1]. У статті ми зосередимось на задачах класифікації в *GIP*-проблемі. В той же час у багатьох прикладних задачах природним чином з'являються матриці в якості математичних «представників» об'єктів аналізованої предметної області. Прикладами такого «матричного подання» є, зокрема, матриці спектрограм в аналізі мовних сигналів та матриці зображень в задачі обробки зображень. Тому природним і нагальним — зважаючи на важливість відповідних прикладних задач — є розвинення засобів опису основних структур для евклідових просторів ширшого, ніж $R^n$ класу, зокрема, для евклідових $(R^{m \times n}, (\cdot, \cdot))$ просторів матриць фіксованої розмірності $m \times n$ з «покоординатними» лінійними операціями та «покоординатним» (слідовим) скалярним добутком. Розширення апарату аналізу на новий простір вимагатиме і вибору слушного класу лінійних операторів та визначення елементів розширеної «матричної алгебри». Як виявляється, таким «адекватним» поставленій задачі операторами виявилися так звані «кортежні оператори» [2, 3], операції з якими визначають розширений варіант «матричної алгебри». Дослідження таких операторів дозволяє в математичному плані отримати конструктивне представлення для SVD-розкладу та побудувати на основі такого представлення теорію ПдО із всіма можливостями застосування оперування основними множинними структурами матричних просторів, аналогічними можливостям для $R^n$ із збереженням конструктивності та складності побудови об'єктів на рівні відповідників для $R^n$ [4]. Таке розширення можливостей оперування дозволило перенести на матриці алгоритми розв'язання *GIP*-задач [3]. Основну інформацію про концепцію «кортежності» та її реалізацію для матричних кортежів наведено в роботі [2]. Нижче наведено необхідну інформацію з цієї роботи. Основну увагу приділено проблемі визначення критеріїв відповідності класам-кластерам в задачах класифікації на основі «занурення» в природні алгебраїчні структури евклідового простору. «Занурення» розуміється як асоціація-породження такої структури наявною послідовністю елементів простору: навчальною вибіркою кластеру в прикладних задачах. Власне, це означає, що такі структури ототожнюються з класами — кластерами, яким можуть належати спостережувані вектори ознак. Проблема занурення розглядається як для векторних, так і для матричних векторів ознак. Для векторних її було розвинуто в роботі [5], для матричних — необхідну теоретичну базу було розвинуто в роботі [6], див. також [7, 8]. В останньому випадку основою побудови відповідного апарату SVD та ПдО є так звані кортежні оператори, які мають і самостійний інтерес. Що стосується самих базових структур, то такими базовими структурами, байдуже: для векторного чи матричного варіанту, що конструктивно описуються засобами ПдО, є лінійні та нелінійні. До перших відносяться лінійні підпростори та гіперплощини, під якими розумітимемо «зміщені підпростори». До других відносимо еліпсоїди (еліпсоїди групування), точніше — еліпсоїдальні циліндри. Реалізація концепції кортежних операторів для матричних кортежів дозволяє побудувати конструктивну теорію ПдО за Муром-Пенроузом, яка уможливлює отримання результатів для матричних евклідових просторів, аналогічних відповідникам для $R^n$. Використання техніки ПдО для кортежних операторів дозволяє постановити та успішно розв'язати для матриць задачі лінійної робастної дискримінації у вигляді [2, 9] для $R^n$. Крім того, наведено результати застосування техніки «занурення» як для $R^n$, так і для $R^{m\times n}$.

## Навчальні вибірки з елементів евклідових просторів та матриці Грама

Вихідна інформація про об'єкти групування в задачах класифікації із учителем міститься в навчальній вибірці, яку позначимо $LS : a_j \in R^m$, $j = \overline{1,n}$ для $R^n$ чи $LS : A_k \in R^{m\times n}, k = \overline{1,K}$ для $R^{m\times n}$. Такі навчальні вибірки пов'язуються з кожним класом-кластером $Kl$, присутнім в GIP-задачі.

Матрицею Грама її елементів називатимемо матрицю, що позначатимемо $G$ та стандартним чином визначатимемо як матрицю скалярних добутків елементів навчальної вибірки

$$G = \left[ \begin{array}{ll} ((a_i, a_j))_{i,j=\overline{1,n}} & \text{— числові вектори,} \\ ((A_i, A_j))_{i,j=\overline{1,K}} & \text{— матриці.} \end{array} \right.$$

Ранг цієї матриці, що співпадає із кількістю лінійно незалежних елементів навчальної вибірки, позначаємо через $r : rankG = r$.

Через $\left(\nu_i, \lambda_i^2\right), i = \overline{1,r}$ позначаємо набір ненульових сингулярностей: власний вектор – власне число, — вважаючи, що сингулярності впорядковані за спаданням власних чисел

$$\lambda_1^2 \geq \cdots \geq \lambda_r^2 > 0.$$

Нагадаймо, що власні вектори представляють собою ортонормований набір векторів

$$(\nu_i, \nu_j) \equiv \nu_i^T \nu_j = \delta_{ij}, \quad i, j = \overline{1,r}.$$

## Засоби опису в структуруванні задачі групування

Фундаментальним матричним об'єктом, який є основою конструктивного математичного апарату оперування в просторі числових векторів, є матриця $A$, складена із елементів навчальної вибірки як із стовпчиків

$$A = (a_1, \ldots, a_n).$$

Зазначимо, що в такому поданні вона може розглядатися як векторний кортеж

$$\alpha = (a_1, \ldots, a_n), a_j \in R^m, j = \overline{1,n},$$

тобто, впорядкований набір елементів певної множини. Вона визначає лінійний оператор, що позначається тою самою літерою $A : R^n \to R^m$, що визначається стандартним чином: через операцію множення матричної алгебри. Нескладно переконатися, що визначений за матричним множенням оператор задовольняє співвідношенню

$$Ax = \sum_{j=\overline{1,n}} x_j a_j, \qquad x = \begin{pmatrix} x_1 \\ \cdots \\ x_n \end{pmatrix} \in R^n. \tag{1}$$

Оскільки за такого опису лінійний оператор пов'язується з векторним кортежем $\alpha = (a_1, \ldots, a_n)$, його можна назвати кортежним оператором, породженим цим кортежем.

Оператор, що визначається для матричної навчальної вибірки заміною векторних елементів $a_j, j = \overline{1,n}$ вибірки на матричні $A_j, j = \overline{1,K}$, з відповідною зміною розмірності числового вектору $x$, відіграє для матричного простору ту саму фундаментальну роль. Він називається кортежним, тому що визначається матричним кортежем $\alpha$, складеним із елементів навчальної вибірки: $\alpha = (A_1, \ldots, A_K)$.

**Означення 1.** Кортежним оператором $\wp_\alpha : R^K \to R^{m \times n}$, що визначається матричним кортежем $\alpha = (A_1, \ldots, A_K)$, будемо називати оператор, який визначається співвідношенням

$$\wp_\alpha x = \sum_{j=\overline{1,K}} x_j A_j, \qquad x = \begin{pmatrix} x_1 \\ \cdots \\ x_K \end{pmatrix} \in R^K. \tag{2}$$

Зауважимо, що співвідношення (1), (2) засвідчують однакову природу операторів $A, \wp_\alpha$ між різними евклідовими просторами із-за їхнього однакового визначення через складові кортежу.

Співвідношення (1), (2) є однаковими з точністю до заміни векторних елементів навчальної вибірки на матричні. Кортежність у назві оператора за (2) визначається тим, що він не може бути реалізований засобами матричного множення, як аналогічний з (1) у випадку лінійного оператора між евклідовими просторами числових векторів, коли він реалізується матрицею оператора.

Позначимо через $\nu_{\alpha i}$, $i = \overline{1,r}$ ортонормований набір «ненульових» власних векторів матриці Грама: власних векторів, що відповідають ненульовим власним числам $\lambda_1^2 \geq \cdots \geq \lambda_r^2 > 0$, $r = rankG$.

Позначимо через $U_{\alpha i}, i = \overline{1,r}$ нормовані $\lambda_i, i = \overline{1,r}$ образи власних векторів матриці Грама за відображенням (1) чи (2) відповідно до типу елементів навчальної вибірки

$$U_{\alpha i} = \left[ \begin{array}{l} \dfrac{1}{\lambda_i} A\nu_i, \\ \dfrac{1}{\lambda_i} \wp_\alpha \nu_i, \end{array} \right. \quad i = \overline{1,r}. \tag{3}$$

В подальшому використовуватимуться центровані середніми елементи навчальної вибірки

$$\begin{array}{ll} \tilde{a}_j = a_j - \bar{a}, j = \overline{1,n}, & \bar{a} = n^{-1} \sum\limits_{j=\overline{1,n}} a_j, \\ \tilde{A}_j = A_j - \bar{a}, j = \overline{1,K}, & \bar{a} = K^{-1} \sum\limits_{j=\overline{1,K}} A_j \end{array} \tag{4}$$

та відповідні кортежі $\tilde{\alpha}$.

Відповідно, використовуватимуться позначення $\tilde{G}$, $\left(\tilde{\nu}_{\tilde{\alpha} i}, \tilde{\lambda}_i^2\right)$, $U_{\tilde{\alpha} i}$, $i = \overline{1,r}$.

Зауважимо, що в позначеннях (4) позначення $\bar{a}$ є однаковим і для середнього за векторним і для середнього за матричним варіантами вибірки. Зміст: векторний чи матричний — легко визначатиметься контекстом.

**Лема 1.** *Набір $U_{\alpha i}$, $i = \overline{1,r}$ , як і $U_{\tilde{\alpha} i}$, $i = \overline{1,r}$, є ортонормованими наборами векторів у відповідному евклідовому просторі: $R^m$ — для векторного варіанту навчальної вибірки, $R^{m \times n}$ — для матричного.*

В подальшому не виділятимуться спеціально результати для кортежів з центрованих елементів навчальної вибірки.

**Лема 2.** *Елементи набору $U_{\alpha i}$, $i = \overline{1,r}$ є розв'язками оптимізаційної задачі побудови нормованих векторів з найбільшими сумами квадратів ортогональних проекцій на них елементів навчальної вибірки з послідовним* *звуженням множини оптимізації до ортогонального доповнення лінійного підпростору, породженого вже знайденими розв'язками. Оптимізаційні значення відповідних сум складають* $\lambda_i^2$, $i = \overline{1,r}$

$$U_{\alpha 1} = \left[ \begin{array}{l} \underset{U:\|U\|=1}{\arg} \max \sum_{j=\overline{1,n}} \|\mathrm{Pr}_U\, a_j\|^2 - \textit{векторна вибірка,} \\ \underset{U:\|U\|=1}{\arg} \max \sum_{j=\overline{1,n}} \|\mathrm{Pr}_U\, A_j\|^2 - \textit{матрична вибірка,} \end{array} \right.$$

$$U_{\alpha i} = \left[ \begin{array}{l} \underset{U:\|U\|=1, U\perp L(U_{\alpha 1},\dots,U_{\alpha\, i-1})}{\arg} \max \sum_{j=\overline{1,n}} \|\mathrm{Pr}_U\, a_j\|^2 - \textit{векторна вибірка,} \\ \underset{U:\|U\|=1, U\perp L(U_{\alpha 1},\dots,U_{\alpha\, i-1})}{\arg} \max \sum_{j=\overline{1,n}} \|\mathrm{Pr}_U\, A_j\|^2 - \textit{матрична вибірка,} \end{array} \right.$$

$i = \overline{2,r}$.

Для елементу $e \in E$ абстрактного евклідового простору $(E, (\cdot,\cdot))$ через $\ell_e$ позначатимемо лінійний функціонал, що породжується цим елементом через скалярний добуток цього простору: $\ell_e z = (z, e)$, $z \in E$.

Визначимо оператори $Z_U$ та $Z_{\tilde{U}}$ в $R^{m\times n}$ (матрична послідовність) чи $R^m$ (векторна послідовність) відповідно, співвідношеннями

$$\begin{aligned} Z_U &= I - \sum_{j=\overline{1,r}} U_{\alpha i} \ell_{U_{\alpha i}}, \\ Z_{\tilde{U}} &= I - \sum_{j=\overline{1,r}} U_{\tilde{\alpha} i} \ell_{U_{\tilde{\alpha} i}}, \end{aligned} \tag{5}$$

де $I$ — тотожний оператор у відповідному просторі.

Нескладно переконатися, що для векторних навчальних вибірок, $Z_\alpha$, може бути записаний у вигляді

$$Z_\alpha = I_m - \sum_{j=\overline{1,r}} U_{\alpha i} U_{\alpha i}^T,$$

де $I_m$ — одинична матриця розмірності $m$.

Аналогічно визначатимуться $Z$-оператори за $Z_v$ та $Z_{\tilde{v}}$ в $R^K$ (чи $R^n$) в залежності від типу елементів навчальної вибірки: матричного чи векторного

$$Z_\nu = I - \sum_{j=\overline{1,r}} \nu_{\alpha i} \nu_{U_{\alpha i}}^T;$$

$$Z_{\tilde{\nu}} = I - \sum_{j=\overline{1,r}} U_{\tilde{\alpha} i} \nu_{\tilde{\nu}_{\alpha i}}^T.$$

**Лема 3.** *Оператори* $I - Z_\alpha, Z_\alpha$ *є операторами ортогонального проектування, відповідно, на множину можливих значень (рейндж) та ортогональне доповнення до нього для операторів* $A, \wp_\alpha$.

**Лема 4.** *Квадрат* $\rho^2$ *евклідової відстані елементу* $Y$ *від лінійних структур* $L_{LS}, \Gamma_{LS}$ *визначається співвідношеннями*

— $\rho^2(Y, L_{LS}) = (Y, Z_{\alpha LS}Y)\rho^2(Y, \Gamma_{LS}) = (Y - \underbrace{a_{LS}}_{\bar{a},\bar{A}}, Z_{\tilde{\alpha}LS}(Y - \underbrace{a_{LS}}_{\bar{a},\bar{A}}))$,

— $\rho^2(Y, \Gamma_{LS}) = (Y - \underbrace{a_{LS}}_{\bar{a},\bar{A}}, Z_{\tilde{\alpha}LS}(Y - \underbrace{a_{LS}}_{\bar{a},\bar{A}}))$.

Визначимо квадратичні форми $R_\alpha$ та $R_{\tilde{\alpha}}$ ($R$-форми) співвідношеннями

$$R_\alpha(U) = \sum_{j=\overline{1,r}} \lambda_i^{-2} \left(\ell_{U_{\alpha i}}, U\right)^2, U \in \left[ \begin{array}{l} R^m\text{— векторна вибірка,} \\ R^{m\times n}\text{— матрична вибірка,} \end{array} \right.$$

$$R_{\tilde{\alpha}}(U) = \sum_{j=\overline{1,r}} \tilde{\lambda}_i^{-2} \left(\ell_{U_{\tilde{\alpha} i}}, U\right)^2, U \in \left[ \begin{array}{l} R^m\text{— векторна вибірка,} \\ R^{m\times n}\text{— матрична вибірка.} \end{array} \right.$$

Позначимо через $r_{\min}, \tilde{r}_{\min}$ максимальні значення квадратичних форм, відповідно $R_\alpha(U)$, $R_{\tilde{\alpha}}(U)$ на вихідних та центрованих елементах навчальної вибірки, байдуже — у векторному чи матричному випадку. Зауважимо, що $r_{\min}, \tilde{r}_{\min}$ не перевищує $r$.

**Теорема 1.** *Для кожного з елементів навчальної вибірки $V$ є справедливими нерівності*

$$r_{\min}^{-1} R_\alpha(V) \le 1, \quad \tilde{r}_{\min}^{-1} R_{\tilde{\alpha}}(V - \bar{a}) \le 1,$$
$$\text{де } V = \left[ \begin{array}{l} a_j, j = \overline{1,n}\text{— векторна вибірка,} \\ A_j, j = \overline{1,n}\text{— матрична вибірка.} \end{array} \right. \tag{6}$$

*Зауважимо, що у співвідношеннях* (6) $r_{\min}, \tilde{r}_{\min}$ *можна замінити на* $r$

$$r^{-1} R_\alpha(V) \le 1, \quad r^{-1} R_{\tilde{\alpha}}(V - \bar{a}) \le 1,$$
$$\text{де } V = \left[ \begin{array}{l} a_j, j = \overline{1,n}\text{— векторна вибірка,} \\ A_j, j = \overline{1,n}\text{— матрична вибірка.} \end{array} \right. \tag{7}$$

Співвідношення (6), (7) свідчать, що елементи навчальної вибірки належать еліпсоїдальним циліндрами:
— центральним (з центром в нулі) з рівняннями

$$r_{\min}^{-1} R_\alpha(U) \le 1,\ r^{-1} R_\alpha(U) \le 1: \quad U \in \left[ \begin{array}{l} R^m\text{— векторна вибірка,} \\ R^{m\times n}\text{— матрична вибірка,} \end{array} \right. \tag{8}$$

$$r_{\min}^{-1} R_\alpha(U) \le 1;$$

— чи нецентральними: з центром в $\bar{a}$ і, відповідно рівняннями

$$r^{-1} R_{\tilde{\alpha}}(U - \bar{a}) \le 1,\ \tilde{r}_{\min}^{-1} R_{\tilde{\alpha}}(U - \bar{a}) \le 1. \tag{9}$$

Осі та довжини напівосей цих еліпсоїдальних циліндрів складають пари відповідно

$$\left(U_{\alpha i}, r^{-1}\lambda_i\right), i = \overline{1,r}, \left(U_{\alpha i}, r_{\min}^{-1}\lambda_i\right), i = \overline{1,r} \text{ — для центральних та}$$

$$\left(U_{\tilde{\alpha} i}, r^{-1}\tilde{\lambda}_i\right), i = \overline{1,r}, \left(U_{\tilde{\alpha} i}, \tilde{r}_{\min}^{-1}\tilde{\lambda}_i\right), i = \overline{1,r} \text{ — для нецентральних.}$$

Еліпсоїдальні циліндри (8), (9) називатимемо еліпсоїдами групування: центральними та нецентральними відповідно.

Еліпсоїди, в описі яких вживається $r_{\min}^{-1}$, $\tilde{r}_{\min}^{-1}$ називатимуться мінімальними: центральними чи нецентральними відповідно.

### Ступінь відповідності — відстані відповідності

У класифікаційному варіанті *GIP*-задач віднесення до кожного із класів-кластерів $Kl$ із множини можливих $\mathbb{C}\ell$ здійснюється за «найкращим» варіантом відповідності класу-кластеру. Звичайним є задання «типового» представника кластеру $p_{Kl}$ та ступеня відповідності у вигляді «відстані» відповідності класу-кластеру: $\rho(Y, Kl), Kl \in \mathbb{C}\ell$. Такою відстанню відповідності часто є евклідова відстань до $p_{Kl}$ — представника кластера

$$\rho(Y, Kl) \equiv \rho(Y, p_{Kl}) = \|Y - p_{Kl}\|, Kl \in \mathbb{C}\ell.$$

Загалом, "ступінь відповідності" може визначатися і специфікованим до кожного кластеру-класу функціоналом

$$\rho(Y, Kl) \equiv \rho_{Kl}(Y, p_{Kl}), Kl \in \mathbb{C}\ell.$$

Віднесення класифікованого елементу $Y$ до класу-кластеру $Kl_r$ здійснюється за мінімальним значенням відстаней до кластерів

$$Kl_r = \arg\min_{Kl \in \mathbb{C}\ell} \rho_{Kl}(Y, p_{Kl}). \tag{10}$$

Такі специфіковані за кластерами-класами функціонали визначають через асоціацію
(«занурення») із кожним кластером тої чи іншої базової структури евклідового простору: лінійної (лінійного підпростору чи гіперплощини: зміщеного простору), — чи нелінійної: еліпсу групування (центрального чи нецентрального, звичайного чи мінімального).

### Відстані відповідності: занурення у лінійні структури

За такого варіанту «занурення» класифікація визначається асоціацією з кожним класом-кластером $Kl \in \mathbb{C}\ell$, представленим навчальною вибіркою.

Відповідно, стандартними базовими лінійними структурами «занурення» є лінійні підпростори $L_{Kl}$ чи гіперплощини $\Gamma_{Kl}$, $Kl \in \mathbb{C}\ell$

$$L_{Kl} = \left[ \begin{array}{l} L(a_j \in R^m, j = \overline{1,n}), \\ L(A_k \in R^{m\times n}, k = \overline{1,K}), \end{array} \right.$$

$$\Gamma_{Kl} = \left[ \begin{array}{l} \Gamma(\bar{a}_{Kl}, L(\tilde{a}_j \in R^m, j = \overline{1,n}) : \ \bar{a}_{Kl} = \frac{1}{n}\sum_{i=1}^n a_i, \tilde{a}_j = a_j - \bar{a}, j = \overline{1,n}, \\ \Gamma(\bar{A}_{Kl}, L(\tilde{A}_k \in R^{m\times n}, k = \overline{1,K}) : \ \bar{A}_{Kl} = \frac{1}{K}\sum_{k=1}^K A_k, \tilde{A}_k = A_k - \bar{A}. \end{array} \right.$$

Представниками $p_{Kl}, Kl \in \mathbb{C}\ell$ кожного з кластерів у випадку лінійних структур є

— $p_{Kl} \equiv 0, Kl \in \mathbb{C}\ell$ для підпросторів,

— $p_{Kl} = \left[ \begin{array}{l} \bar{a}_{Kl} \text{ — векторна вибірка} \\ \bar{A}_{Kl} \text{ — матрична вибірка} \end{array} \right.$ класу $Kl \in \mathbb{C}\ell$ для гіперплощин.

Відстанями відповідності до представників — евклідові відстані до відповідно підпросторів чи гіперплощин «занурення»

— $\rho^2_{Kl}(Y, p_{Kl}) \equiv \rho^2(Y, L_{Kl}), Kl \in \mathbb{C}\ell$ для занурення у лінійні підпростори,

— $\rho^2_{Kl}(Y, p_{Kl}) \equiv \rho^2(Y, \Gamma_{Kl}), Kl \in \mathbb{C}\ell$ для занурення у гіперплощини.

**Теорема 2.** *Квадрат $\rho^2$ відстані відповідності за "занурення" у лінійну структуру представляється співвідношеннями, відповідно:*

$$\rho^2_{Kl}(Y, p_{Kl}) = \rho^2\left(Y, \underbrace{Kl}_{L_{Kl}, \Gamma_{Kl}}\right) =$$

$$= \left[\begin{array}{ll} ((Y, Z_{\alpha Kl} Y)) & \text{— підпростір,} \\ \left(Y - \underbrace{p_{Kl}}_{\bar{a}_{Kl}, \bar{A}_{Kl}}, Z_{\tilde{\alpha} Kl}(Y - \underbrace{p_{Kl}}_{\bar{a}_{Kl}, \bar{A}_{Kl}})\right) & \text{— гіперплощина,} \end{array}\right. \tag{11}$$

$$Kl \in \mathbb{C}\ell.$$

Відстані відповідності: занурення у квадратичні структури

Занурення у квадратичні структури відповідає асоціації з кожним з кластерів еліпсу групування: для всіх кластерів одного і того самого типу — центрального чи нецентрального, звичайного чи мінімального. Представниками еліпсів є

— $p_{Kl} \equiv 0, Kl \in \mathbb{C}\ell$ для центральних,

— $p_{Kl} = \left[\begin{array}{l} \bar{a}_{Kl} \text{ — векторна вибірка} \\ \bar{A}_{Kl} \text{ — матрична вибірка} \end{array}\right.$ класу $Kl \in \mathbb{C}\ell$ для нецентральних.

Відстань $\rho\,(Y, Kl)$ до кластеру визначається як мінімальне значення лінії рівня центрального чи нецентрального, мінімального чи звичайного еліпса групування, що містить елемент $Y$, який піддається класифікації.

**Теорема 3.** *Квадрат $\rho^2\,(Y, Kl)\,, Kl \in \mathbb{C}\ell$ визначається співвідношеннями*

$$\rho^2(Y, Kl) \equiv \rho^2_{Kl}(Y, p_{Kl}) =$$

$$= \left[\begin{array}{ll} \left\{\begin{array}{l} r^{-1} R_{\alpha Kl}(Y) \text{ — звичайний} \\ r^{-1}_{\min} R_{\alpha Kl}(Y) \text{ — мінімальний} \end{array}\right. & \text{— центральні еліпсоїди,} \\ \left\{\begin{array}{l} r^{-1} R_{\tilde{\alpha} Kl}(Y - p_{Kl}) \text{ — звичайний} \\ r^{-1}_{\min} R_{\tilde{\alpha} Kl}(Y - p_{Kl}) \text{ — мінімальний} \end{array}\right. & \text{— нецентральні еліпсоїди,} \end{array}\right.$$

$$\rho^2\,(Y, Kl) \equiv \rho^2_{Kl}(Y, p_{Kl}) =$$

$$= \left[\begin{array}{ll} (Y, Z_{\alpha Kl} Y)) & \text{— підпростір,} \\ \left(Y - \underbrace{p_{Kl}}_{\bar{a}_{Kl}, \bar{A}_{Kl}}, Z_{\tilde{\alpha} Kl}(Y - \underbrace{p_{Kl}}_{\bar{a}_{Kl}, \bar{A}_{Kl}})\right) & \text{— гіперплощина,} \end{array}\right.$$

$$Kl \in \mathbb{C}\ell.$$

### SVD-подання операторів та процедура фільтрації

SVD-подання стандартних лінійних операторів, представлених матрицями [1], чи кортежних матриць [2, 3] має вигляд

$$\left.\begin{matrix} A \\ \wp_\alpha \end{matrix}\right] = \sum_{j=\overline{1,r}} \lambda_i U_{\alpha i} v_{\alpha i}^T.$$

Як зазначалося вище, концепція «кортежних операторів» дозволяє перенесення всіх результатів щодо базових структур в евклідових просторах числових векторів на випадок евклідових просторів матриць фіксованої розмірності. Наявність засобів оперування з базовими структурами в матричних просторах суттєво розширює можливості використання таких просторів в математичному моделюванні, оскільки в багатьох важливих у прикладному відношенні задачах математичного моделювання саме матриці з'являються як представники аналізованих об'єктів — власне в ролі «матричних векторів ознак». Такими важливими сферами застосування можуть бути задачі розпізнавання аудіосигналів, де природним матричним представником слова є його спектрограма, та аналіз зображень, де зображення із самого початку є матрицею. Процедура SVD-фільтрації полягає у використанні SVD-представлення для перетворення набору векторів, що представляють один і той самий об'єкт, як це має місце чи в навчальній вибірці одного класу-кластеру, чи у спектрограмі аудіо сигналу, чи у випадку зображення. Перетворення набору векторів, що представляють навчальну вибірку, пропонується здійснювати через відкидання в SVD-розкладі матриці, для якої вектори є елементами стовпчикового представлення, старших членів з наступним поверненням до векторів стовпчикового представлення перетвореної матриці. Так само, відкиданням старших членів SVD-розкладу матриці здійснюється перетворення матриці, коли вона із самого початку є представником об'єкту. Результат процедури SVD-фільтрації на прикладі мовних сигналів «дерево» та «звук», представлено на Рис. 1.

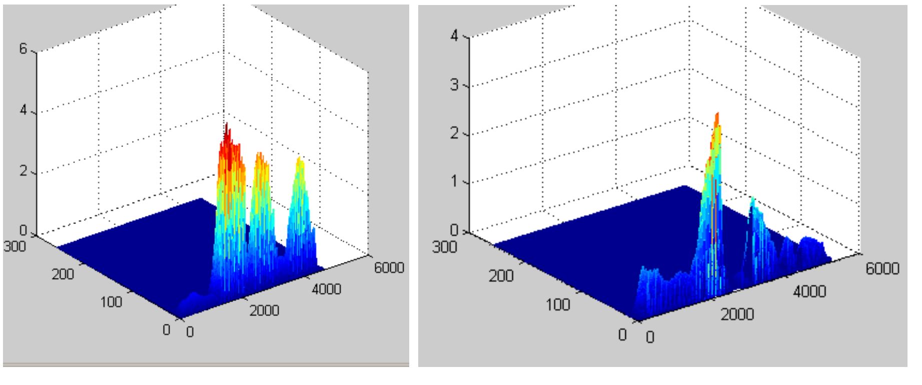


Рис 1. Результат SVD-фільтрації слів «дерево» та «звук»

## Література

1. Kirichenko N. F. Analytical representation of perturbation of pseudoinverse matrices // Cybernetics and system analysis.— 1997. — Vol. 33, №2. — P. 230–239.
2. Донченко В. С., Зинько Т. П. Концепция кортежности и ее реализация для матричных кортежей // Международный научно-технический журнал «Проблемы управления и информатики». — 2016. — №4. — С. 87–99.
3. Donchenko V., Zinko T., Skotarenko F. «Feature vectors» in grouping information problem in applied mathematics: vectors and matrixes // Problems of computer Intellectualization. ITHEA, Kiyv, Ukraine-Sofia, Bulgaria. — 2012. — P. 111–124.
4. Донченко В., Кривонос Ю., Омардибирова В. Базовые структуры евклидовых пространств: конструктивные методы описания и использования // New Trends in Classification and Data Mining. — ITHEA, Sofia, Bulgaria. — 2010. — P. 155–170.
5. Кириченко Н. Ф. Аналитическое представление возмущений псевдообратных матриц // Кибернетика и системный анализ. — 1997. — №2. — С. 98–107.
6. Donchenko V., Nazaraga I., Tarasova O. Matrixes least squares method and examples of its application // International Journal Information Technologies & Knowledge. — 2013. — Vol. 7, №4. — P. 325–336.
7. Донченко В. С. Евклидовы пространства: конструктивные методы описания базовых структур и их использование // Information Models of Knowledge. — Editors: Krassimir Markov, Vitalii Velichko, Oleksy Voloshin. — ITHEA. — Kiev, Ukraine — Sofia, Bulgaria. — 2010. — №19. — P. 362–376.
8. Kirichenko N. F., Donchenko V. S., Serbaev D. P. Nonlinear recursive nonlinear transformation: dynamic systems and optimization // Cybernetics and system analysis. — 2005. — Vol. 41, №3. — P. 364–373.
9. Голік А. О., Донченко В. С. Застосування еліпсоїдальної та ортогональної відстаней відповідності у задачах розпізнавання мови та жестів. // Вісник Київського національного університету імені Тараса Шевченка. Серія: фіз.-мат. науки. — 2013. — Вип. 1. — С. 146–155.

Надійшла 01.10.2016