УДК 681.518:004.93.1’

***Г. А. СТАДНИК***

# СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ, ЩО ФУНКЦІОНУЄ В РЕЖИМІ АВТОМАТИЧНОЇ КЛАСИФІКАЦІЇ

Розглядається інформаційно-екстремальний алгоритм функціонування системи підтримки прийняття рішень в режимі автоматичної класифікації. При цьому запропонований алгоритм дозволяє на етапі самонавчання системи автоматично формувати вхідну навчальну матрицю, а на етапі екзамену, тобто безпосередньо в робочому режимі, виділяти новий клас розпізнавання і здійснювати донавчання системи. Реалізацію запропонованого алгоритму здійснено на прикладі діагностування опортуністичних інфекцій у ВІЛ-інфікованих осіб.

**Ключові слова:** система підтримки прийняття рішень, автоматична класифікація, факторний кластер-аналіз, оптимізація, інформаційний критерій, опортуністична інфекція.

Рассматривается информационно-экстремальный алгоритм функционирования системы поддержки принятия решений в режиме автоматической классификации. При этом предложенный алгоритм позволяет на этапе самообучения системы автоматически формировать входную обучающую матрицу, а на этапе экзамена, то есть непосредственно в рабочем режиме, выделять новый класс распознавания и осуществлять дообучение системы. Реализация предложенного алгоритма выполнена на примере диагностирования оппортунистических инфекций у ВИЧ-инфицированных лиц.

**Ключевые слова:** система поддержки принятия решений, автоматическая классификация, факторный кластер-анализ, оптимизация, информационный критерий, оппортунистическая инфекция.

The process of input data automation classification is considered in the framework of information-extreme intelligence technology, which is based on maximizing the information capacity of the recognition system in the process of its machine learning. The proposed algorithm allows forming the input matrix in the mode of system self-learning and allocation of a new pattern in the mode of system examination. This system operation algorithm implements the modified k-means method for binary Hamming space in the mode of cluster analysis and forms the set of unrecognized patterns in the mode of factor cluster analysis, which is based on statistical criteria for the stability and homogeneity. If representativeness condition of formed set is fulfilled in the exam mode, it is added to the input matrix. The resulting retraining system is implemented for advanced classes’ recognition alphabet. Optimization of the decision support system learning parameters is carried out by searching global maximum of the Kullback information criterion, which is calculated in the working (acceptable) domain of its function. Implementation of the proposed algorithm is executed by the example of opportunistic infections diagnosis in HIV-infected people.

**Keywords:** decision support system, cluster analysis, factor cluster analysis, optimization, information criterion, opportunistic infection.

**Вступ.** Основним шляхом підвищення функціональної ефективності системи підтримки прийняття рішень (СППР) є надання їй властивості адаптивності на основі машинного самонавчання та розпізнавання образів. Розв’язання цієї задачі на практиці здійснюється шляхом розробки алгоритмів автоматичної класифікації [1–3]. При цьому основні науково-методологічні труднощі, що виникають під час інформаційного синтезу здатних самонавчатися СППР, обумовлені довільними початковими умовами формування образів та їх перетином у просторі ознак розпізнавання. Один із перспективних напрямів інформаційного аналізу і синтезу СППР, що функціонує в режимі автоматичної класифікації, полягає в застосуванні ідей і методів інформаційно-екстремальної інтелектуальної технології (ІЕІ-технологія) аналізу даних, що ґрунтується на максимізації інформаційної спроможності системи в процесі її машинного самонавчання [4–6]. При цьому в рамках ІЕІ-технології автоматична класифікація дозволяє розв’язувати дві основні задачі: автоматично формувати вхідний математичного опис здатної самонавчатися СППР (кластер-аналіз вхідних даних) і в режимі екзамену виділяти нові класи розпізнавання, що характеризують можливі функціональні стани об’єкту дослідження (факторний кластер-аналіз).

У статті розглядається інформаційно-екстремальний алгоритм самонавчання СППР, що функціонує в режимах кластер-аналізу і факторного кластер-аналізу (ФКА) даних, отриманих при діагностуванні опортуністичних інфекцій у ВІЛ-інфікованих осіб.

**Постановка задачі.** Розглянемо формалізовану постановку задачі інформаційного синтезу здатної самонавчатися СППР для діагностування опортуністичних інфекцій у ВІЛ-інфікованих осіб, яка функціонує в режимі автоматичної класифікації.

Нехай відома некласифікована багатовимірна навчальна матриця $||y_i^{(j)}|i=\overline{1,N},\ j=\overline{1,n}||$, де $N$, $n$ – кількість ознак розпізнавання і реалізацій образів відповідно. При цьому

$$N = N_1 + N_2,$$

де $N_1$ – кількість дійсних ознак розпізнавання, одержаних за результатами клініко-лабораторних та імуногенетичних досліджень;

$N_2$ – кількість бінарних ознак розпізнавання, отриманих за результатами анамнезу.

Дано вектор параметрів навчання СППР

$$g = < x_m, d_m, \delta >, \qquad (1)$$

де $x_m$ – еталонний вектор-реалізація класу $X_m^{\circ}$, $m=\overline{1,M}$;

$d_m$ – радіус гіперсферичного контейнера класу $X_m^{\circ}$, що відновлюється в радіальному базисі простору ознак розпізнавання;

$\delta$ – параметр поля контрольних допусків на ознаки розпізнавання.

*ISSN 2079-0023 (print)*
*ISSN 2410-2857 (online)*

При цьому задано такі обмеження: $x_m$ – вектор, вершина якого визначає геометричний центр контейнера класу $X_m^{\text{o}}$; $d_m \in [0; d(x_m \oplus x_c) - 1]$, де $d(x_m \oplus x_c)$ – міжцентрова кодова відстань для класу $X_m^{\text{o}}$ і найближчого до нього класу $X_c$ і параметр поля контрольних допусків $\delta \in [0; \delta_{\text{H}} / 2]$, де $\delta_{\text{H}}$ – нормоване поле допусків для відносної шкали вимірювання ознак, яке є областю значень для параметра $\delta$ поля контрольних допусків на ознаки розпізнавання.

Необхідно в процесі самонавчання СППР трансформувати шляхом допустимих перетворень вхідну некласифіковану навчальну матрицю у нечітку класифіковану і побудувати чітке розбиття класів розпізнавання $\Re^{|M|}$, де $M$ – задана кількість класів розпізнавання, які характеризують функціональні стани досліджуваного процесу. Для цього визначити оптимальні значення координат вектору параметрів навчання (1), які забезпечують максимальне значення усередненого за алфавітом класів розпізнавання інформаційного критерію функціональної ефективності (КФЕ) машинного самонавчання системи

$$\overline{E}^{*} = \frac{1}{M} \sum_{m=1}^{M} \max_{\{k\}} E_m^{(k)}, \qquad (2)$$

де $E_m^{(k)}$ – інформаційний КФЕ навчання СППР, значення якого обчислюються на $k$-му кроці самонавчання;

$\{k\}$ – впорядкована множина кроків навчання.

В режимі екзамену СППР повинна розв'язувати такі задачі:

1) прийняття рішень про належність реалізації образу, що розпізнається, до одного із класів сформованого на етапі самонавчання алфавіту $\{X_m^{\text{o}}\}$;

2) формування з нерозпізнаних реалізацій образів додаткової навчальної матриці, яка за умови її репрезентативності приєднується до вхідної навчальної матриці і здійснюється перенавчання СППР для розширеного алфавіту класів розпізнавання $\{X_m^{\text{o}}\}^{|\Lambda|}$, де $\Lambda$ – символ відкритості множини.

**Функціонування СППР у режимі інформаційно-екстремального кластер-аналізу вхідних даних.** Використання кластер-аналізу в методах ІЕІ-технології дозволяє автоматизувати процес формування вхідної нечіткої класифікованої навчальної матриці $||y_{m,i}^{(j)}||$, допустимі перетворення якої в субпарацептуальному дискретному просторі ознак дозволяють побудувати в процесі машинного навчання чітке його розбиття на класи розпізнавання. Таким чином, процес навчання в рамках ІЕІ-технології можна розглядати як процедуру дефазіфікації нечітких даних без застосування функції належності Заде, яка по суті є аналогом функції щільності ймовірностей.

Реалізація інформаційно-екстремального кластер-аналізу вхідних даних будемо здійснювати шляхом включення модифікованого для бінарного простору Хеммінга методу *k*-середніх у контур інформаційно-екстремального алгоритму навчання СППР. При цьому згідно з концепцією ІЕІ-технології початковий розподіл векторів-реалізацій в евклідовому просторі ознак розпізнавання в процесі кластер-аналізу трансформується в бінарний простір Хеммінга шляхом пошуку оптимальної системи контрольних допусків на ознаки розпізнавання, що забезпечує максимальне значення інформаційного критерію (2).

Як початкові еталонні вектори кластерів, контейнери яких відновлюються в радіальному просторі ознак розпізнавання, обираються реалізації некласифікованої бінарної навчальної матриці $\{x_i^{(j)} \mid i = \overline{1, N},\ j = \overline{1, n}\}$, найбільш віддалені одна від одної, тобто найближчі до вершин відповідно нульового та одиничного векторів, а початкові значення інших $M - 2$ центрів кластерів визначаються шляхом рівномірного поділу їх міжцентрової відстані на $M - 1$ відрізків. При цьому запропонований підхід дозволяє усунути такий недолік методу *k*-середніх, як чутливість до вибору початкових значень центрів кластерів, та забезпечує виконання принципу максимальної різноманітності між реалізаціями різних класів розпізнавання.

Необхідною умовою інформаційно-екстремального синтезу СППР, що функціонує в режимі кластер-аналізу вхідних даних, є виконання обмежень [4]

$$[\forall X_m^{\text{o}} \in \tilde{\Re}^{|M|}]\{X_m^{\text{o}} \neq \varnothing\}, \qquad (3)$$

$$[\forall X_m^{\text{o}} \in \tilde{\Re}^{|M|}][\forall X_c^{\text{o}} \in \tilde{\Re}^{|M|}]\{X_m^{\text{o}} \neq X_c^{\text{o}} \rightarrow$$

$$\rightarrow X_m^{\text{o}} \cap X_c^{\text{o}} \neq \varnothing \mid\ m, c = \overline{1, M}\}, \qquad (4)$$

$$[\exists X_m^{\text{o}} \in \tilde{\Re}^{|M|}][\exists X_c^{\text{o}} \in \tilde{\Re}^{|M|}]\{X_m^{\text{o}} \neq X_c^{\text{o}} \rightarrow$$

$$\rightarrow KerX_m^{\text{o}} \cap KerX_c^{\text{o}} = \varnothing\}, \qquad (5)$$

де $\varnothing$ – символ порожньої множини;

$KerX_m^{\text{o}}$, $KerX_c^{\text{o}}$ – ядра класу $X_m^{\text{o}}$ і найближчого до нього класу $X_c^{\text{o}}$ відповідно, які визначаються в бінарному просторі вершинами їх двійкових еталонних векторів;

$$[\forall X_m^{\text{o}} \in \tilde{\Re}^{|M|}][\forall X_c^{\text{o}} \in \tilde{\Re}^{|M|}]\ \{X_m^{\text{o}} \neq X_c^{\text{o}} \rightarrow$$

$$\rightarrow (d_m^{*} < d(KerX_m^{\text{o}} \oplus KerX_c^{\text{o}})) \wedge$$

$$\wedge (d_c^{*} < d(KerX_m^{\text{o}} \oplus KerX_c^{\text{o}}))\}, \qquad (6)$$

де $d_m^{*}$, $d_c^{*}$ – оптимальні радіуси гіперсферичних контейнерів класів $X_m^{\text{o}}$ і $X_c^{\text{o}}$ відповідно;

$d(KerX_m^{\text{o}} \oplus KerX_c^{\text{o}})$ – міжцентрова кодова відстань кластерів $X_m^{\text{o}}$ і $X_c^{\text{o}}$;

$$\bigcup_{X_m^o \in \widetilde{\Re}^{|M|}} X_m^o \subseteq \Omega_Б, \qquad (7)$$

де $\Omega_Б$ – бінарний простір ознак розпізнавання.

Категорійну модель СППР, що функціонує в режимі інформаційно-екстремального кластер-аналізу вхідних даних, показано на рис. 1.

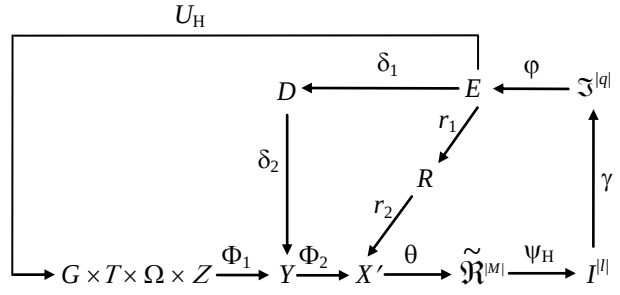



Рис. 1 – Категорійна модель інформаційно-екстремального кластер-аналізу вхідних даних

Категорійна модель (рис. 1), яка відображає узагальнену структуру алгоритму самонавчання СППР, містить оператор формування вхідного математичного опису $\Phi_1 : G \times T \times Z \times \Omega \to Y$, де $G$ – простір вхідних сигналів (факторів), які діють на СППР; $T$ – множина моментів часу зняття інформації; $Z$ – простір можливих функціональних станів СППР; $\Omega$ – простір ознак розпізнавання; $Y$ – вибіркова множина на вході СППР. Оператор $\Phi_2 : Y \to X'$ формує некласифіковану бінарну навчальну матрицю. Оператор $\theta$ здійснює відображення навчальної матриці $X'$ на розбиття $\widetilde{\Re}^{|M|}$ в двійковому просторі Хеммінга шляхом агрегації двійкових векторів-реалізацій, що знаходяться в межах поточного радіусу гіперсферичного контейнера. Оператор класифікації $\Psi_Н : \widetilde{\Re} \to I^{|l|}$ перевіряє основну статистичну гіпотезу про належність реалізації $\{x_i^{(j)} \mid i = \overline{1,N},\ j = \overline{1,n}\}$ нечіткому кластеру $X_m^o$, де $I^{|l|}$ – множина $l$ статистичних гіпотез. Оператор $\gamma : I^{|l|} \to \Im^{|q|}$ шляхом оцінки результатів статистичної перевірки гіпотез формує множину точнісних характеристик $\Im^{|q|}$, де $q = l^2$ – кількість точнісних характеристик. Оператор $\varphi : \Im^{|q|} \to E$ обчислює множину $E$ значень інформаційного КФЕ, який є функціоналом точнісних характеристик. Оператор $r_1 : E \to R$ оптимізує геометричні параметри нечіткого розбиття $\widetilde{\Re}^{|M|}$ шляхом пошуку максимуму КФЕ навчання СППР розпізнавати реалізації кластеру $X_m^o$ на множині $R$ допустимих радіусів контейнера кластеру $X_m^o$. Оператор $r_2 : R \to X'$ регламентує центрування та збільшення радіусу гіперсферичного контейнера кластеру $X_m^o$. Центрування контейнера кластеру в бінарному просторі здійснюється за принципом $k$-середніх і триває до моменту порушення умови (6).

Оптимізація системи контрольних допусків $D$ на ознаки розпізнавання здійснюється за багатоциклічною ітераційною процедурою, в якій послідовно реалізуються оператори $\Phi_2$, $\theta$, $\Psi$, Н, $\gamma$, $\varphi$ і оператори $\delta_1$, $\delta_2$ цілеспрямованої зміни множини $D$. Оператор $U_Н : E \to G \times T \times \Omega \times Z$ регламентує процес навчання СППР.

Алгоритм інформаційно-екстремального кластер-аналізу даних у рамках ІЕІ-технології подамо як двоциклічну ітераційну процедуру пошуку глобального максимуму інформаційного КФЕ (2) в робочій (допустимій) області визначення його функції

$$\delta^* = \arg \max_{G_\delta} \{ \max_{G_E \cap \{k\}} \overline{E}^{(k)} \}, \qquad (8)$$

де $\overline{E}^{(k)}$ – усереднене за алфавітом класів розпізнавання значення інформаційного критерію, обчислене на $k$-му кроці кластеризації;

$G_\delta$ – допустима область значень параметра $\delta$ поля контрольних допусків;

$G_E$ – робоча область визначення функції критерію $\overline{E}$;

$G_d$ – допустима область значень радіуса гіперсферичного контейнера класу розпізнавання.

Вхідні дані для алгоритму інформаційно-екстремального кластер-аналізу даних: кількість класів розбиття $M$, параметр $\delta$ поля контрольних допусків на ознаки розпізнавання; масив реалізацій образу $\{y_i^{(j)} \mid i = \overline{1,N},\ j = \overline{1,n}\}$.

Розглянемо основні кроки реалізації алгоритму за умови розбиття простору ознак розпізнавання на $M$ кластерів. При цьому формування класифікованої навчальної матриці $\{y_{m,i}^{(j)} \mid m = \overline{1,M}, i = \overline{1,N}, j = \overline{1,n_m}\}$ будемо здійснювати за умови, що кожним відновленим кластером досягнений заданий репрезентативний обсяг векторів-реалізацій $n_m \geq n_{\min}$.

1. Ініціалізація лічильника кроків зміни параметра $\delta$: $l := 0$.

2. $l := l + 1$;

3. Обчислення нижнього $A_{НК_i}[l]$ та верхнього $A_{ВК_i}[l]$ контрольних допусків для $i$-ї ознаки розпізнавання

$$A_{НК_i}[l] = y_i - \delta_i \frac{\delta_{Н_i}}{100};\ A_{ВК_i}[l] = y_i + \delta_i \frac{\delta_{Н_i}}{100}, \qquad (9)$$

де $y_i$ – $i$-та ознака еталонного вектору-реалізації $y$ масиву реалізацій образу $\{y_i^{(j)} \mid i = \overline{1,N_1},\ j = \overline{1,n}\}$.

4. Формування бінарної навчальної матриці $\{x_i^{(j)} \mid i = \overline{1,N},\ j = \overline{1,n}\}$ за правилом

$$x_i^{(j)} = \begin{cases} 1, \text{ if } \{A_{НК_i}[l] < y_i^{(j)} < A_{ВК_i}[l]\} \wedge y_i^{(j)} \in \{S_k\}; \\ 0, \text{ if } \{A_{ВК_i}[l] \geq y_i^{(j)} \vee y_i^{(j)} \leq A_{НК_i}[l]\} \wedge y_i^{(j)} \in \{S_k\}; \\ y_i^{(j)}, \text{ if } y_i^{(j)} \notin \{S_k\}, \end{cases}$$

*ISSN 2079-0023 (print)*
*ISSN 2410-2857 (online)*

де $A_{\text{НК}_i}[l]$ і $A_{\text{ВК}_i}[l]$ – нижній і верхній контрольні допуски для *i*-ї ознаки еталонного вектору *y* масиву реалізацій образу $\{y_i^{(j)} \mid i=\overline{1,N_1},\ j=\overline{1,n}\}$, обчислені за формулами (9);

$y_i^{(j)}$ – значення *i*-ї ознаки в *j*-й реалізації масиву реалізацій образу $\{y_i^{(j)} \mid i=\overline{1,N},\ j=\overline{1,n}\}$;

$\{S_k \mid k=\overline{1,N_1}\}$ – множина дійсних ознак розпізнавання.

5. Знаходження початкової множини $\{x_m\}$ еталонних векторів класів $\{X_m^{\text{o}}\}$ за умови, що

$$< x_1, x_2 >:= \arg\max_{i,j} d(x^{(i)} \oplus x^{(j)}).$$

6. Ініціалізація радіусів контейнерів класу $X_m^{\text{o}}$:

$$d_m := d(x_1 \oplus x_2) - 1,\ m=\overline{1,M}.$$

7. Ініціалізація лічильника прогонів процедури алгоритму самонавчання СППР: $s := 0$ і початкового усередненого значення максимуму КФЕ самонавчання СППР: $\overline{E}^{(s)} := 0$.

8. $s := s+1$.

9. Ініціалізація лічильника класів розпізнавання: $m := 0$.

10. $m := m+1$.

11. Ініціалізація лічильника кроків зміни радіуса контейнера кластеру $X_m^{\text{o}}$: $d_m := 0$.

12. $d_m := d_m + 1$.

13. Ініціалізація масиву $A_{n_m}$ значень $n_m$ кластерів $X_m^{\text{o}}$: $A_{n_m} := 0$, $m=\overline{1,M}$.

14. Розбиття на *M* кластерів реалізацій бінарної навчальної матриці $\{x_i^{(j)} \mid i=\overline{1,N},\ j=\overline{1,n}\}$ за умови, що

$$d(x_m \oplus x^{(j)}) \le d_m,$$

де $d(x_m \oplus x^{(j)})$ – кодова відстань Хеммінга між двійковими векторами $x_m$ і $x^{(j)}$ та формування нечіткої класифікованої бінарної навчальної матриці $\{x_{m,i}^{(j)} \mid m=\overline{1,M},\ i=\overline{1,N},\ j=\overline{1,n_m}\}$.

15. Додавання в масив $A_{n_m}$ поточних значень $n_m$ кластерів $X_m^{\text{o}}$: $A_{n_m} := n_m$, $m=\overline{1,M}$.

16. Обчислення інформаційного КФЕ самонавчання СППР.

17. Формування еталонного вектору $x_m$ кластеру $X_m^{\text{o}}$ за правилом

$$x_{m,i} = \begin{cases} 1, \text{ if } \dfrac{1}{n_m}\sum_{j=1}^{n_m} x_{m,i}^{(j)} > 0{,}5; \\ 0, \text{ if else}, \end{cases}$$

де $x_{m,i}^{(j)}$ – значення *i*-ї ознаки в *j*-й реалізації бінарної навчальної матриці $\{x_{m,i}^{(j)} \mid m=\overline{1,M},\ i=\overline{1,N},\ j=\overline{1,n_m}\}$ кластеру $X_m^{\text{o}}$.

18. Порівняння: якщо $d_m < d(x_1 \oplus x_2)$, то виконується пункт 12, інакше – пункт 19.

19. Знаходження максимуму КФЕ в робочій області його визначення:

$$E_m^{(s)} := \max_{\{d_m\}} E_m,$$

за умови, що $n_m \ge n_{\min}$, і визначення оптимальних еталонного вектору та радіусу кластеру $X_m^{\text{o}}$:

$$< x_m^{(s)}, d_m^{(s)} >:= \arg\max_{d_m} E_m.$$

При цьому у випадку відсутності робочої області визначення КФЕ як квазіоптимальні еталонний вектор та радіус кластеру $X_m^{\text{o}}$ приймають їх початкові значення.

20. Порівняння: якщо $m < M$, то виконується пункт 10, інакше – пункт 21.

21. Порівняння: якщо $\overline{E}^{(s)} > \overline{E}^{(s-1)}$, де $\overline{E}^{(s)}$, $\overline{E}^{(s-1)}$ – усереднені значення максимумів КФЕ, обчислені за формулою (2) відповідно на $s$-му і $s-1$-му прогонах процедури алгоритму самонавчання СППР, то

$$< \hat{x}_m^{(s)}, \hat{d}_m^{(s)} >:=< x_m^{(s)}, d_m^{(s)} >$$

і виконується пункт 8, інакше

$$\overline{E}^{(l)} := \overline{E}^{(s)},$$

$$< \hat{x}_m^{(l)}, \hat{d}_m^{(l)} >:=< x_m^{(s)}, d_m^{(s)} >,$$

та виконується пункт 22.

22. Порівняння: якщо $\delta_i \le \delta_{\text{H}_i}/2$, то виконується пункт 2, інакше – пункт 23.

23. Виконується процедура пошуку глобального максимуму КФЕ

$$\overline{E}^{*} := \max_{\{\delta\}} \overline{E}^{(l)}$$

в робочій області визначення його функції і оптимальних еталонного вектору та радіусу кластеру $X_m^{\text{o}}$

$$< x_m^{*}, d_m^{*} >:= \arg\max_{\{\delta\}} \overline{E}^{(l)}.$$

24. ЗУПИН.

Для реалізацій, які не вдалося розпізнати в режимі екзамену, передбачається в рамках IEI-технології формування за методом ФКА нового кластеру даних.

Як загальний критерій валідації розбиття простору діагностичних ознак на кластери розглядається робоча формула модифікації КФЕ за Кульбаком [7], що дозволяє використовувати під час процедури самонавчання СППР різні за обсягами

навчальні вибірки для класів розпізнавання алфавіту $\{X_m^{\text{o}}\}$

$$E_m^{(k)} = \frac{\{n_c - n_m + 2\cdot(K_{1,m}^{(k)} - K_{2,m}^{(k)})\}}{n_c + n_m} \times$$

$$\times \log_2\left(\frac{n_c + (K_{1,m}^{(k)} - K_{2,m}^{(k)}) + 10^{-r}}{n_m - (K_{1,m}^{(k)} - K_{2,m}^{(k)}) + 10^{-r}}\right), \qquad (10)$$

де $K_{1,m}^{(k)}$, $K_{2,m}^{(k)}$ – кількість подій, що характеризують належність реалізацій образу до контейнера класу $X_m^{\text{o}}$ якщо вони дійсно є реалізаціями відповідно класу $X_m^{\text{o}}$ та найближчого до нього класу $X_c^{\text{o}}$ на *k*-му кроці навчання СППР;

$n_m$, $n_c$ – обсяг навчальної вибірки для класів $X_m^{\text{o}}$ і $X_c^{\text{o}}$ відповідно.

Нормовану модифікацію критерію (10) можна подати у вигляді:

$$\hat{E}_m^{(k)} = \frac{E_m^{(k)}}{E_{m,\max}^{(k)}}, \qquad (11)$$

де $E_m^{(k)}$ – КФЕ, обчислене за формулою (10);

$E_{m,\max}^{(k)}$ – максимальне значення критерію (10), обчислене при значеннях $K_{1,m}^{(k)} = n_m$ та $K_{2,m}^{(k)} = 0$.

Таким чином, алгоритм функціонування СППР, що навчається в режимі кластер-аналізу вхідних даних, в рамках IEI-технології полягає в ітераційній процедурі наближення глобального максимуму інформаційного КФЕ до граничного значення, обчисленого в робочій (допустимій) області визначення його функції.

**Категорійна модель і алгоритм інформаційно-екстремального ФКА.** Категорійну модель СППР, яка функціонує в режимі інформаційно-екстремального ФКА, показано на рис. 2.

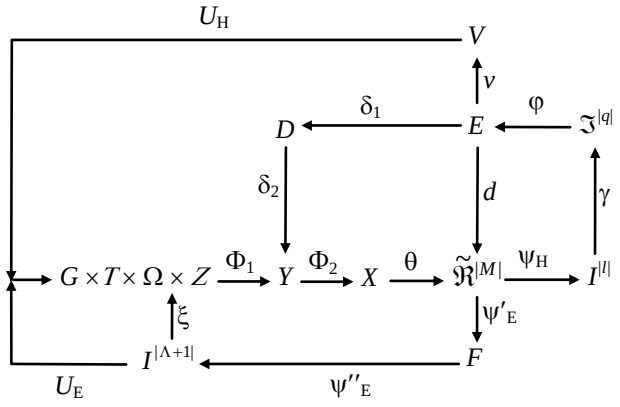



Рис. 2 – Категорійна модель інформаційно-екстремального ФКА

На відміну від діаграми, представленої на рис. 1, оператор $\Phi_2 : Y \to X$ категорійної моделі (рис. 2) формує класифіковану бінарну навчальну матрицю, оператор $\theta : X \to \tilde{\Re}^{|M|}$ відновлює на кожному кроці навчання оптимальне в інформаційному розумінні розбиття простору ознак на $M$ класів розпізнавання, а оператор класифікації $\Psi_{\text{H}} : \tilde{\Re}^{|M|} \to I^{|l|}$ перевіряє основну статистичну гіпотезу про належність реалізації $\{x_{m,i}^{(j)} \mid i = \overline{1,N},\ j = \overline{1,n_m}\}$ класу $X_m^{\text{o}}$. Контур оптимізації геометричних параметрів розбиття $\tilde{\Re}^{|M|}$ шляхом пошуку максимуму КФЕ навчання розпізнаванню реалізацій класу $X_m^{\text{o}}$ замикається оператором $d : E \to \tilde{\Re}^{|M|}$.

Особливість діаграми, показаної на рис. 2, полягає в наявності паралельних контурів самонавчання і екзамену. При цьому оператор класифікації екзаменаційної реалізації утворює композицію $\Psi_{\text{E}} = \Psi'_{\text{E}} \circ \Psi''_{\text{E}}$, де $\Psi'_{\text{E}}$, $\Psi''_{\text{E}}$ – оператори обчислення функції належності реалізації образу контейнеру та реалізації вирішальних правил відповідно. За результатами екзамену формується відкрита множина гіпотез $I^{|\Lambda+1|}$, серед яких гіпотеза $\gamma_{\Lambda+1}$ означає, що екзаменаційна реалізація не належить алфавіту класів розпізнавання $\{X_m^{\text{o}}\}^{|\Lambda|}$. Оператор $I^{|\Lambda+1|} \to Z$ формує новий простір функціональних станів керованого динамічного процесу і запускає контур самонавчання СППР.

Зовнішній контур категорійної моделі (рис. 2) містить множину $V$ типів вирішальних правил, що будуються в радіальному базисі простору ознак розпізнавання. При цьому згідно з принципом відкладених рішень Івахненка О. Г. оператор $\nu : E \to V$ здійснює вибір нового типу вирішальних правил за умови, що після оптимізації параметрів самонавчання значення КФЕ (2) не досягає свого граничного максимального значення.

Оператори категорійної моделі, представленої на рис. 2,

$$U_{\text{H}} : V \to G\times T\times\Omega\times Z,$$

$$U_{\text{E}} : I^{|\Lambda+1|} \to G\times T\times\Omega\times Z,$$

регламентують процеси самонавчання й екзамену СППР відповідно.

**Приклад реалізації інформаційно-екстремальних алгоритмів автоматичної класифікації.** Спочатку розглянемо процес інформаційно-екстремального кластер-аналізу для формування навчальної матриці, для розпізнавання ступеню тяжкості перебігу ВІЛ. Обсяг матриці складав $n = 80$, а розмірність структурованих векторів-реалізацій становила $N = 63$ клініко-лабораторних та імуногенетичних ознак розпізнавання, які відповідно характеризують:

1) загальний стан пацієнта при зверненні за медичною допомогою, ураження органів і систем, показники клінічного, біохімічного аналізу крові;

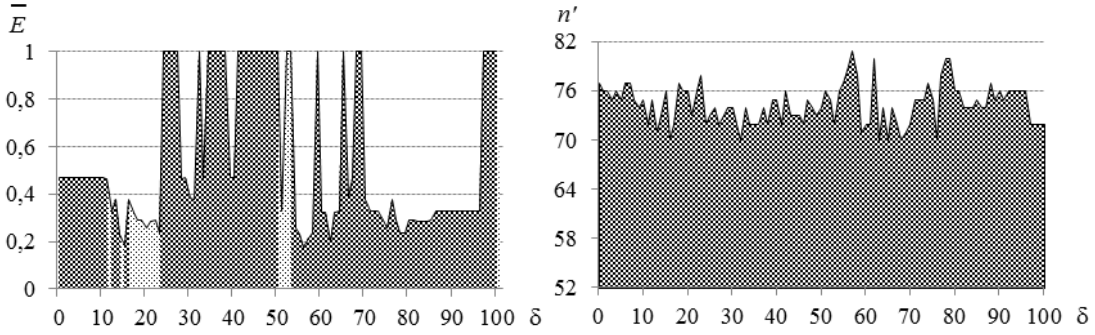
*ISSN 2079-0023 (print)*
*ISSN 2410-2857 (online)*

2) дослідження рівнів популяцій лімфоцитів та сироваткові рівні IL-4, IL-10, TNF-α, поліморфізми поодиноких нуклеотидів генів цитокінів IL-4 (-590C/T), IL-10 (-592C/A), TNF-α (-308G/A).

Процес оптимального розбиття простору діагностичних ознак на кластери розглянемо при $n_{\min} = 35$.

З метою зовнішньої валідації результатів функціонування СППР у режимі кластер-аналізу вхідних даних апостеріорна класифікація навчальної матриці здійснювалася експертами предметної області. В результаті реалізації алгоритму (8) вхідну некласифіковану матрицю було розбито на два класи по 40 і 36 реалізації в кожному. Ці класи характеризували ВІЛ-інфікованих осіб з тяжким (кількість опортуністичних інфекцій на одного хворого більше 3-х) і середньотяжким (кількість опортуністичних інфекцій на одного хворого – 1-2-ві) перебігом захворювання та відповідні функціональні стани патологічного процесу.

На рис. 3, *а* і 3, *б* показано графіки зміни усередненого значення КФЕ самонавчання СППР та кількості реалізацій навчальної вибірки, що потрапили до контейнерів класів розпізнавання в процесі паралельної оптимізації параметра δ поля контрольних допусків на ознаки розпізнавання за алгоритмом (8).

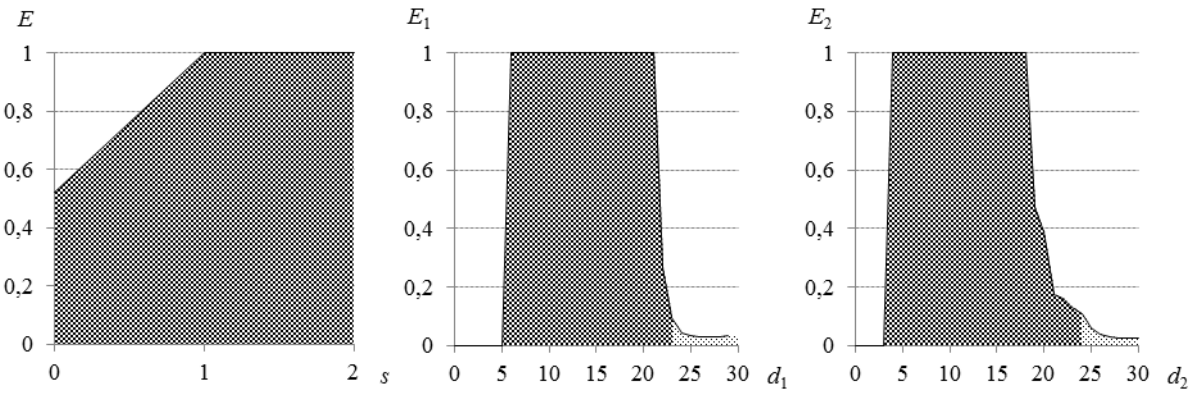



*а* *б*

Рис. 3 – Оптимізація параметра поля контрольних допусків на ознаки розпізнавання:
*а* – графік залежності усередненого за алфавітом класів розпізнавання критерію Кульбака;
*б* – кількість реалізацій навчальної вибірки, що потрапили до контейнерів класів розпізнавання

Аналіз рис. 3 показує, що квазіоптимальне значення параметра поля контрольних допусків дорівнює $\delta^* = \pm 42\%$ від номінального (усередненого) для вхідної некласифікованої навчальної матриці значення ознак розпізнавання при максимальному значенні усередненого нормованого КФЕ $\overline{E}^* = 1$. Таким чином, оптимізація системи контрольних допусків на ознаки розпізнавання дозволяє побудувати безпомилкові за навчальною матрицею вирішальні правила.

На рис. 4 наведено графіки залежності КФЕ (11) від радіусів гіперсферичних контейнерів класів розпізнавання, одержаних після виконання процедури оптимізації контрольних допусків на ознаки розпізнавання за паралельним алгоритмом (8).




*а* *б* *в*

Рис. 4 – Графіки залежності критерію Кульбака від: *а* − кількості прогонів процедури алгоритму самонавчання СППР;
*б* − радіуса контейнера класу розпізнавання $X_1^{\text{o}}$; *в* − радіуса контейнера класу розпізнавання $X_2^{\text{o}}$

Аналіз рис. 4 показує, що максимальне усереднене значення КФЕ самонавчання СППР, отримане на другій ітерації роботи алгоритму, досягає свого максимального граничного значення і дорівнює $\overline{E}^* = 1$. При цьому оптимальні значення радіусів відповідних контейнерів класів розпізнавання

дорівнюють $d_1^* = 21$ і $d_2^* = 16$, а кількість векторів-реалізацій образів, що потрапили до кластерів – $n_1 = 40$ і $n_2 = 36$.

На етапі екзамену, тобто при функціонуванні СППР в режимі ФКА, оцінювалась достовірність розпізнавання 110 реалізацій навчальної вибірки, що характеризує ступінь тяжкості перебігу ВІЛ. При цьому рішення приймалося шляхом визначення максимального значення вирішального правила у вигляді геометричної функції належності, яка для гіперсферичного класифікатора і реалізацій класу $X_m^{\text{o}}$ має вигляд [4]

$$\mu_m = 1 - \frac{d(x_m^* \oplus x_{\text{e}})}{d_m^*}, \qquad (12)$$

де $x_m$ – еталонний вектор контейнера класу розпізнавання $X_m^{\text{o}}$;

$d_m^*$ – радіус гіперсфричного контейнера класу розпізнавання $X_m^{\text{o}}$;

$x_{\text{e}}$ – реалізація навчальної вибірки, що розпізнається.

За результатами фізичного моделювання сформовано розширений алфавіт класів розпізнавання $\{X_m^{\text{o}}\}^{\Lambda}$, до якого увійшли:

1) клас $X_1^{\text{o}}$ (35 реалізацій), що характеризував групу осіб з тяжким перебігом захворювання (кількість опортуністичних інфекцій на одного хворого більше трьох);

2) клас $X_2^{\text{o}}$ (35 реалізацій) – група осіб з середньотяжким перебігом захворювання (кількість опортуністичних інфекцій на одного хворого – одна–дві);

3) клас $X_3^{\text{o}}$ (40 реалізацій) – контрольна група осіб (практично здорові донори крові).

При цьому клас розпізнавання $X_3^{\text{o}}$ сформовано з реалізацій екзаменаційної вибірки, значення функції належності (12) яких дорівнює $\mu_m = -1, m = \overline{1, M}$, тобто з реалізацій, не віднесених до жодного з контейнерів класів розпізнавання, відновлених у режимі самонавчання СППР.

На рис. 5 показано графік зміни усередненого значення КФЕ навчання СППР в процесі оптимізації параметра $\delta$ поля контрольних допусків на ознаки розпізнавання за паралельним алгоритмом, приведеним у праці [8], для розширеного алфавіту класів розпізнавання $\{X_m^{\text{o}}\}^{\Lambda}$.

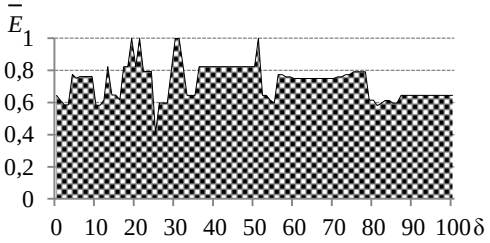
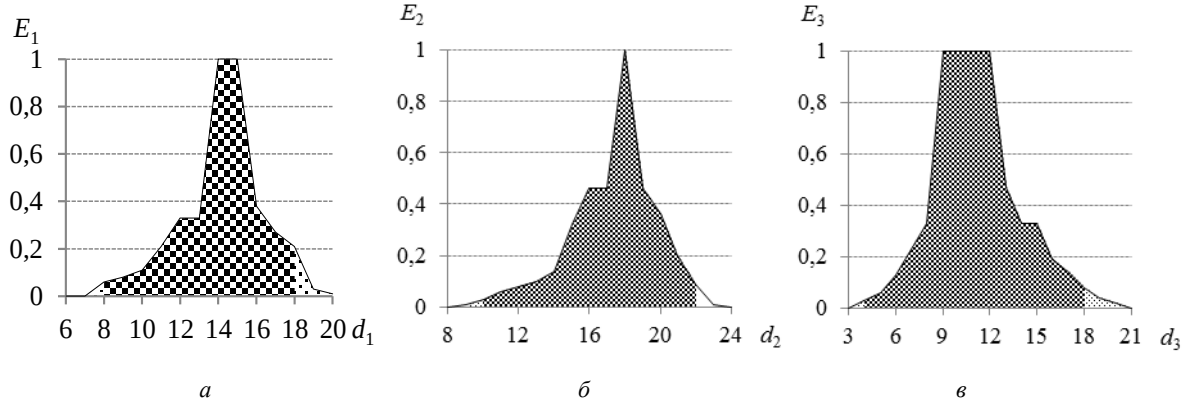



Рис. 5 – Оптимізація параметра поля контрольних допусків на ознаки розпізнавання для алфавіту класів розпізнавання $\{X_m^{\text{o}}\}^{\Lambda}$

Аналіз рис. 5 показує, що оптимальне значення параметра поля контрольних допусків дорівнює $\delta^* = \pm 51\%$ від номінального (усередненого) для базового класу $X_1^{\text{o}}$ значення ознак розпізнавання при максимальному значенні усередненого КФЕ $\overline{E}^* = 1$, що свідчить про побудову безпомилкових за навчальною матрицею вирішальних правил.

На рис. 6 наведено графіки залежності КФЕ (11) від радіусів гіперсферичних контейнерів класів розпізнавання, одержаних при застосуванні оптимального значення параметра $\delta^* = \pm 51\%$ поля контрольних допусків на ознаки розпізнавання. Аналіз рис. 6 показує, що максимальне усереднене значення КФЕ навчання СППР досягає свого граничного значення і дорівнює $\overline{E}^* = 1$, а оптимальні значення радіусів відповідних контейнерів класів розпізнавання, які формують вирішальні правила (12) відповідно – $d_1^* = 14$, $d_2^* = 18$ і $d_3^* = 9$.




*а* *б* *в*

Рис. 6 – Графіки залежності критерію Кульбака від радіусів контейнерів класів розпізнавання з алфавіту $\{X_m^{\text{o}}\}^{\Lambda}$:

*а* – клас $X_1^{\text{o}}$; *б* – клас $X_2^{\text{o}}$; *в* – клас $X_3^{\text{o}}$

Таким чином, за результатами фізичного моделювання підтверджено працездатність і надійність розробленого інформаційного та програмного забезпечення СППР, що функціонує в режимі інформаційно-екстремального факторного кластер-аналізу.

**Висновки.** Запропоновано в рамках інформаційно-екстремального синтезу здатної самонавчатися СППР алгоритм кластер-аналізу вхідних даних, побудований на основі модифікованого для бінарного простору Хеммінга методу *k*-середніх, що дозволило автоматично сформувати вхідну класифіковану навчальну матрицю і шляхом оптимізації параметрів функціонування системи побудувати безпомилкові за навчальною матрицею вирішальні правила. Крім того, запропоновано інформаційно-екстремальний алгоритм факторного кластер-аналізу, що дозволило безпосередньо в робочому режимі виділяти нові класи розпізнавання і здійснювати перенавчання СППР.

**Список літератури**

1. *Duda R. O.* Pattern Classification : second ed. / *R. O. Duda, P. E. Hart, D. G. Stork*. – New York : John Wiley & Sons, 2001. – 738 p.
2. Data Clustering : Algorithms and Applications / [ed. by *C. C. Aggarwal, C. K. Reddy*]. – CRC Press, 2013. – 652 p.
3. Анализ данных и процессов / *[А. А. Барсегян [и др.]*. – [3-е изд.]. – СПб. : БХВ-Петербург, 2009. – 512 с.
4. *Довбиш А. С.* Основи проектування інтелектуальних систем : Навчальний посібник / *А. С. Довбиш*. – Суми : Вид-во СумДУ, 2009. – 171 с.
5. *Москаленко В. В.* Інформаційно-екстремальне навчання системи підтримки прийняття рішень з адаптивною кластеризацією даних / *В. В. Москаленко* // Вісник СумДУ. Серія «Технічні науки». – 2012. – № 3. – С. 110–124.
6. *Стадник Г. А.* Інформаційно-екстремальна кластеризація діагностичних даних / *Г. А. Стадник* // Інтелектуальні системи в промисловості і освіті : Четверта міжн. наук.-практ. конф., 6-8 лист. 2013 р. : тези доповідей. – Суми : СумДУ, 2013. – С. 93–94.
7. *Dovbysh A. S.* Information-Extreme Algorithm for Optimizing Parameters of Hyperellipsoidal Containers of Recognition Classes / *A. S. Dovbysh, N. N. Budnyk, V. V. Moskalenko* // Journal of automation and information sciences. − 2012. – V. 44, I. 10. − P. 35–44.
8. *Довбиш А. С.* Система підтримки прийняття рішень для визначення схеми лікування гострої кишкової інфекції / *А. С. Довбиш, Г. А. Стадник, К. С. Полов'ян* // Вісник СумДУ. Серія «Технічні науки». – 2012. – № 1. – С. 25–31.

**References (transliterated)**

1. Duda R. O., Hart P. E., Stork D. G. *Pattern Classification : second ed.* New York, John Wiley Publ., 2001. 738 p.
2. Aggarwal C. C., Reddy C. K., ed. *Data Clustering : Algorithms and Applications*. CRC Press Publ., 2013. 652 p.
3. Barsegyan A. A., Kupriyanov M. S., Kholod I. I., Tess M. D., Elizarov S. I. *Analiz dannyh i processov* [Analysis of data and processes]. 3rd ed. SPb., BHV-Peterburg Publ., 2009. 512 p.
4. Dovbysh A. S. *Osnovy proektuvannya intelektualnykh system : Navchalnyy posibnyk* [Fundamentals of Intelligent Systems : Tutorial]. Sumy, SumDU Publ., 2009. 171 p. **5.** Moskalenko V. V. Informatsiino-ekstremalne navchannia systemy pidtrymky pryiniattia rishen z adaptyvnoiu klasteryzatsiieiu danykh [Adaptive self-learning of information-extreme decision support system]. *Visnyk SumDU. Series. "Tekhnichni nauky"* [Bulletin of the Sumy State University. Series "Technical sciencies"]. Sumy, 2012, no. 3, pp. 110–124.
6. Stadnyk H. A. Informatsiino-ekstremalna klasteryzatsiia diahnostychnykh danykh [Information extreme clustering of diagnostic data]. *Intelektual'ni systemy v promyslovosti i osviti. Tezy dopovidey IV mizhnarodnoyi naukovo-praktychnoyi konferentsiyi (6–8 lystopada 2013 r., Sumy)* [Intelligence Systems in Industry and Education. Abstracts of the IV Int. Sci.-Pract. Conf. (6–8 Nov. 2013, Sumy)]. Sumy, SumDU Publ., 2013, p. 93–94.
7. Dovbysh A. S., Budnyk N. N., Moskalenko V. V. Information-Extreme Algorithm for Optimizing Parameters of Hyperellipsoidal Containers of Recognition Classes. *Journal of automation and information sciences*. USA, 2012, no. 44, issue 10, pp. 35–44.
8. Dovbysh A. S., Stadnyk H. A., Polovian K. S. Systema pidtrymky pryiniattia rishen dlia vyznachennia skhemy likuvannia hostroi kyshkovoi infektsii [Decision support system for determination of acute enteric infection treatment regimen]. *Visnyk SumDU. Series. "Tekhnichni nauky"* [Bulletin of the Sumy State University. Series "Technical sciencies"]. Sumy, 2012, no. 1, pp. 25–31.

*Надійшла (received) 25.02.2016*

*Бібліографічні описи / Библиографические описания / Bibliographic descriptions*

**Система підтримки прийняття рішень, що функціонує в режимі автоматичної класифікації / Г. А. Стадник** // Вісник НТУ «ХПІ». Серія: Системний аналіз, управління та інформаційні технології. – Х. : НТУ «ХПІ», 2016. – № 37 (1209). – С. 35–42. – Бібліогр.: 8 назв. – ISSN 2079-0023.

**Система поддержки принятия решений, функционирующая в режиме автоматической классификации / А. А. Стадник** // Вісник НТУ «ХПІ». Серія: Системний аналіз, управління та інформаційні технології. – Харків : НТУ «ХПІ», 2016. – № 37 (1209). – С. 35–42. – Библиогр.: 8 назв. – ISSN 2079-0023.

**Decision Support System, functioning in the automatic classification / H. A. Stadnyk** // Bulletin of NTU "KhPI". Series: System analysis, control and information technology. – Kharkiv : NTU "KhPI", 2016. – No. 37 (1209). – P. 35–42. – Bibliogr.: 8. – ISSN 2079-0023.

*Відомості про авторів / Сведения об авторах / About the Authors*

***Стадник Ганна Анатоліївна*** – Сумський державний університет, аспірант кафедри комп'ютерних наук; тел.: (099) 433-53-65; e-mail: anna.stadnyk@gmail.com.

***Стадник Анна Анатольевна*** – Сумский государственный университет, аспирантка кафедры компьютерных наук; тел.: (099) 433-53-65; e-mail: anna.stadnyk@gmail.com.

***Stadnyk Hanna Anatoliivna*** – Sumy State University, Ph. D. Student at the Department of Computer Sciences; tel.: (099) 433-53-65; e-mail: anna.stadnyk@gmail.com.