

УДК 681.518:004.93.1'

А.С. ДОВБИШ, Г.А. СТАДНИК, А.І. ПІДДУБНА

Сумський державний університет, Україна

ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНИЙ АЛГОРИТМ НАВЧАННЯ УНІМОДАЛЬНОГО КЛАСИФІКАТОРА ДЛЯ ДІАГНОСТУВАННЯ ОПОРТУНІСТИЧНИХ ІНФЕКЦІЙ У ВІЛ-ІНФІКОВАНИХ ОСІБ

Пропонується інформаційно-екстремальний алгоритм аналізу і синтезу комп'ютеризованої системи діагностування (КСД) інфекційних патологій, побудованої на базі унімодального класифікатора, який характеризується єдиним центром розсіювання векторів-реалізацій образів. При цьому на етапі навчання системи розглядається задача побудови вирішальних правил шляхом побудови оптимального в інформаційному розумінні розбиття простору ознак на класи розпізнавання з оптимізацією контрольних допусків на діагностичні ознаки розпізнавання за паралельним алгоритмом. За запропонованим алгоритмом розроблено інформаційне та програмне забезпечення системи підтримки прийняття рішень, яка є основною складовою КСД, для діагностування опортуністичних інфекцій у ВІЛ-інфікованих осіб.

Ключові слова: інформаційно-екстремальний алгоритм, комп'ютеризована система діагностування, унімодальний класифікатор, навчання, контрольні допуски, функціональна ефективність, екзамен, опортуністична інфекція.

Вступ

Підвищення функціональної ефективності діагностування при лікуванні інфекційних патологій за умов апріорної невизначеності пов'язано з розробкою та впровадженням інтелектуальних інформаційних технологій, що базуються на ідеях і методах машинного навчання та розпізнавання образів. Відомі методи розпізнавання образів [1, 2] в основному носять модельний характер, оскільки вони не враховують перетин класів розпізнавання, довільні умови формування навчальних матриць та вимагають використання навчальних вибірок великих обсягів, що робить неефективним їх використання у практичних задачах діагностування та лікування.

Для підвищення функціональної ефективності комп'ютеризованих систем діагностування (КСД) інфекційних патологій перспективним є використання ідей і методів інформаційно-екстремальної інтелектуальної технології (ІЕІ-технології), яка ґрунтується на максимізації інформаційної спроможності системи підтримки прийняття рішень (СППР), яка є основною складовою КСД у процесі її навчання [3-5].

У статті [6] розглядався алгоритм інформаційно-екстремального навчання діагностичної СППР для діагностування інфекційних патологій з застосуванням мультимодального класифікатора, що характеризується наявністю декількох центрів розсіювання реалізацій класів із заданого алфавіту. Але у багатьох задачах медичного діагностування існує потреба застосування унімодального класифікатора

з єдиним загальним центром розсіювання та впорядкованим алфавітом класів розпізнавання, наприклад, при діагностуванні із трьохальтернативною системою оцінок типу «Норма», «Менше норми», «Більше норми».

У статті розглядається алгоритм інформаційно-екстремального навчання діагностичної СППР для діагностування опортуністичних інфекцій у ВІЛ-інфікованих осіб за паралельним алгоритмом оптимізації контрольних допусків на діагностичні ознаки розпізнавання.

1. Постановка задачі

Розглянемо КСД, складовою частиною якої є здатна навчатися система підтримки прийняття рішень (СППР). Нехай дано $\{X_m^0 \mid m = \overline{1, M}\}$ – структурований алфавіт M класів розпізнавання, що характеризують відповідні функціональні стани патологічного процесу і навчальну багатовимірну матрицю типу «об'єкт-властивість»

$$\|y_{m,i}^j, \mid i = \overline{1, N}, j = \overline{1, n}\|,$$

де N, n – кількість діагностичних ознак розпізнавання і векторів-реалізацій образу відповідно, x_m – вектор, вершина якого визначає єдиний геометричний центр розсіювання векторів-реалізацій всіх класів в бінарному просторі ознак Ω_B . Відомо структурований вектор параметрів функціонування СППР $g = \langle d_m, \delta \rangle$, де d_m – радіус контейнера класу X_m^0 ,

що відновлюється в радіальному базисі дискретного простору ознак; δ – параметр поля системи контрольних допусків (СКД) на діагностичні ознаки розпізнавання. При цьому задано такі обмеження: радіуси контейнерів класів розпізнавання, що відновлюються в радіальному базисі бінарного простору ознак, знаходяться у відношенні порядку $d_m > d_{m-1}$, а радіус найвіддаленішого контейнера від центру класу X_M^0 дорівнює $d_M = N$.

Необхідно в процесі навчання СППР визначити оптимальні значення координат вектора g , які забезпечують максимум усередненого за алфавітом інформаційного критерію функціональної ефективності (КФЕ) навчання СППР

$$\bar{E}^* = \frac{1}{M} \sum_{m=1}^M \max_{\{k\}} E_m, \quad (1)$$

де E_m – інформаційний КФЕ навчання СППР розпізнавати реалізації класу X_m^0 ; $\{k\}$ – впорядкована множина кроків навчання.

У режимі екзамену необхідно з максимальною достовірністю прийняти рішення про належність реалізації образу, що розпізнається, до одного із класів алфавіту $\{X_m^0\}$.

2. Алгоритм навчання СППР

Алгоритм навчання СППР у рамках ІЕІ-технології подамо як двоциклічну ітераційну процедуру пошуку глобального максимуму інформаційного КФЕ (1) в робочій (допустимій) області визначення його функції

$$\delta^* = \arg \max_{G_\delta} \{ \max_{G_E \cap G_d} \bar{E} \}, \quad (2)$$

де G_δ – допустима область значень параметра δ поля контрольних допусків; G_E – робоча область визначення функцій критерію \bar{E} ; G_d – допустима область значень радіуса гіперсферичного контейнера класу розпізнавання.

Розглянемо алгоритм паралельної оптимізації контрольних допусків на діагностичні ознаки розпізнавання, який відновлює в процесі навчання оптимальні контейнери класів розпізнавання. При цьому відносно базового класу X_1^0 визначаються за процедурою (2) оптимальні контрольні допуски при їх одночасній зміні для всіх неперервних ознак розпізнавання. Вхідні дані: масив реалізацій образу $\|y_{m,i}^{(j)}\|$, $i = \overline{1, N}$, $j = \overline{1, n}$; система нормованих допусків $\{\delta_{H,i}\}$, яка визначає область значень відповід-

них контрольних допусків. За область значень параметра δ приймається інтервал $[0; \delta_H / 2]$.

Розглянемо основні етапи реалізації алгоритму:

1. Обнуляється лічильник кроків зміни параметра δ : $l := 0$.

2. Запускається лічильник: $l := l + 1$ і обчислюються нижні $A_{HK_i}[l]$ та верхні $A_{BK_i}[l]$ контрольні допуски для всіх неперервних ознак:

$$A_{HK_i}[l] = y_{1,i} - \delta \frac{\delta_{H,i}}{100}; \quad A_{BK_i}[l] = y_{1,i} + \delta \frac{\delta_{H,i}}{100}, \quad (3)$$

де $y_{1,i}$ – i -та ознака еталонного вектора-реалізації y_1 базового класу X_1^0 , який характеризує найбільш бажаний для особи, що приймає рішення, функціональний стан СППР.

3. Формується бінарна навчальна матриця $\|x_{m,i}^{(j)}\|$ за правилом

$$x_{m,i}^{(j)} = \begin{cases} 1, & \text{if } \{A_{HK_i}[l] < y_{m,i}^{(j)} < A_{BK_i}[l]\} \wedge x_{m,i}^{(j)} \in R; \\ 0, & \text{if } \{A_{BK_i}[l] \geq y_{m,i}^{(j)} \vee y_{m,i}^{(j)} \leq A_{HK_i}[l]\} \wedge \\ & \wedge x_{m,i}^{(j)} \in R; \\ y_{m,i}^{(j)}, & \text{else.} \end{cases}$$

4. Формується еталонний вектор-реалізація x класів розпізнавання за правилом

$$x_i = \begin{cases} 1, & \text{if } \frac{1}{n} \sum_{j=1}^n x_i^{(j)} > 0,5; \\ 0, & \text{else.} \end{cases}$$

5. Для кожного класу розпізнавання відновлюється контейнер за схемою:

а) обнуляється лічильник класів розпізнавання: $m := 0$;

б) $m := m + 1$;

в) обнуляється лічильник кроків зміни радіуса контейнера: $d := 0$;

г) $d := d + 1$;

д) обчислюється значення нормованого інформаційного КФЕ (1);

е) якщо $E_m \notin G_E$, то виконується пункт бг, інакше – пункт бє;

є) виконується процедура пошуку глобального максимуму КФЕ (1) в робочій області визначення його функції;

ж) $E_m^*[l] := \underset{\{d\}}{\text{extrem}} E_m[l, d]$;

з) визначається оптимальний радіус контейнера

класу X_m^0 : $d_m^*[l] := \arg \text{extrem}_{\{d\}} E_m[l, d]$.

6. Якщо $m \leq M$, то виконується пункт 5б, інакше – пункт 7.

7. Обчислюється за формулою (1) усереднений КФЕ $\bar{E}[l]$.

8. Якщо $\delta[l] \leq \delta_H / 2$, то виконується пункт 2, інакше – пункт 9.

9. Якщо $\bar{E}[l] \notin G_E$, то виконується пункт 10, інакше – пункт 2.

10. Виконується процедура пошуку глобального максимуму КФЕ $\bar{E}^*[l] := \text{extrem}_{\{l\}} E_m[l]$ в робочій області визначення його функції;

11. Визначається оптимальний параметр поля контрольних допусків на діагностичні ознаки розпізнавання: $\delta^* := \arg \bar{E}^*[l]$ і за формулою (3) обчислюються оптимальні контрольні допуски на діагностичні ознаки розпізнавання $A_{HK_i}^*$ та $A_{BK_i}^*$.

12. ЗУПИН

Як критерій оптимізації параметрів навчання СППР у рамках ІЕІ-технології розглянемо інформаційну міру Кульбака, яка є функціоналом від точнісних характеристик. Робоча модифікація міри Кульбака для унімодального класифікатора має вигляд [3,5]

$$E_m^{(k)} = \frac{1}{n} \log_2 \left(\frac{n + (K_{1,m}^{(k)} - K_{2,m}^{(k)}) + 10^{-r}}{n - (K_{1,m}^{(k)} - K_{2,m}^{(k)}) + 10^{-r}} \right) * (K_{1,m}^{(k)} - K_{2,m}^{(k)}), \quad (4)$$

де $K_{1,m}^{(k)}$ – кількість подій, що визначають належність реалізацій образу контейнеру класу X_m^0 , якщо дійсно вони є реалізаціями цього класу; $K_{2,m}^{(k)}$ – кількість подій, що визначають належність реалізацій образу контейнеру класу X_m^0 , якщо дійсно вони є реалізаціями іншому класу; n – обсяг репрезентативної навчальної вибірки; 10^{-r} – число цифр в мантісі значення критерію $E_m^{(k)}$.

Визначення належності реалізації $x^{(j)}$ контейнеру класу X_m^0 для унімодального класифікатора здійснюється за правилом

$$\text{if } d_{m-1} < d[x \oplus x^{(j)}] < d_m \text{ then} \\ x^{(j)} \in X_m^0; \text{ else } x^{(j)} \notin X_m^0,$$

де d_{m-1} – визначений в процесі навчання оптимальний радіус контейнера внутрішнього (вкладеного) класу (при цьому для класу X_1^0 внутрішній радіус $d_0 = 0$); $d[x \oplus x^{(j)}]$ – кодова відстань вектора $x^{(j)}$ від центра розсіювання реалізацій, що визначається вершиною усередненого вектора x ; \oplus – символ операції складання за модулем два; d_m – поточний радіус контейнера класу X_m^0 , що відновлюється в бінарному просторі ознак розпізнавання.

Нормовану модифікацію критерію (5) можна подати у вигляді:

$$\hat{E}_m^{(k)} = \frac{E_m^{(k)}}{E_{m,\max}^{(k)}}, \quad (5)$$

де $E_m^{(k)}$ – КФЕ, обчислене за формулою (4); $E_{m,\max}^{(k)}$ – максимальне значення критерію (4), обчислене при значеннях $K_{1,m}^{(k)} = n$ та $K_{2,m}^{(k)} = 0$.

3. Приклад реалізації алгоритму навчання

Розглянемо застосування вищевказаного алгоритму навчання СППР за трьох альтернативною системою оцінок: «Норма», «Менше норми» і «Більше норми» для діагностування опортуністичних інфекцій у ВІЛ-інфікованих осіб. Алфавіт класів розпізнавання складався із трьох класів. При цьому клас X_1^0 характеризує контрольну групу осіб (практично здорові донори крові). Клас X_2^0 – ВІЛ-інфіковані особи з середньо тяжким перебігом захворювання (кількість опортуністичних інфекцій на одного хворого – 1-2-і) і клас X_3^0 – ВІЛ-інфіковані особи з тяжким перебігом захворювання (кількість опортуністичних інфекцій на одного хворого більше 3-х). Навчальні матриці класів мали по 35 реалізацій, кожна з яких складалася з 63 ознак розпізнавання. При цьому вектори-реалізації класів подано у вигляді структурованої послідовності клініко-лабораторних та імуногенетичних ознак розпізнавання, які характеризують: загальний стан пацієнта при зверненні за медичною допомогою, ураження органів і систем, показники клінічного, біохімічного аналізу крові, дослідження рівнів популяцій лімфоцитів та сироваткові рівні ІЛ-4, ІЛ-10, TNF- α , поліморфізми поодиноких нуклеотидів генів цитокінів ІЛ-4 (-590C/T), ІЛ-10 (-592C/A), TNF- α (-308G/A).

На рис. 1 показано графік залежності усередненого КФЕ від параметра поля допусків δ , одержан-

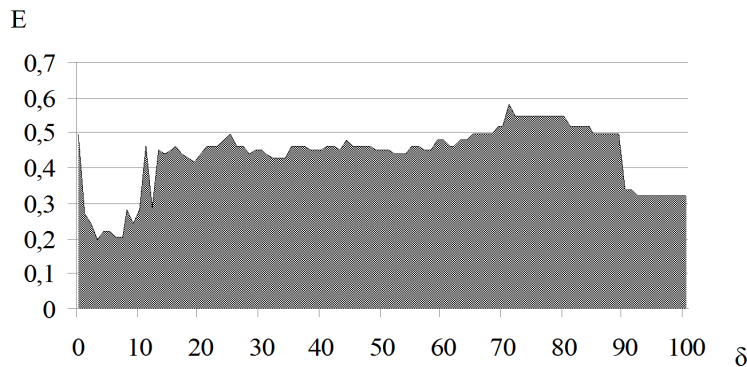


Рис. 1. Графік залежності критерію Кульбака від величини параметра поля контрольних допусків для базового класу X_1^0

ний в процесі паралельної оптимізації системи контрольних допусків на діагностичні ознаки розпізнавання. Тут і далі темними ділянками позначено робочі (допустимі) області визначення функції критерію (5), в яких значення достовірностей перевищують відповідні помилки першого та другого роду і одночасно виконується умова $d_m > d_{m-1}$.

Аналіз рис. 1 показує, що оптимальне значення параметра поля контрольних допусків дорівнює $\delta^* = \pm 71\%$ від усереднених для класу X_1^0 значень ознак розпізнавання при усередненому нормованому максимальному значенні КФЕ $\bar{E}^* = 0,58$.

На рис. 2 показано графіки залежності КФЕ (5) від радіусів гіперсферичних контейнерів класів розпізнавання, одержаних при застосуванні системи оптимальних контрольних допусків на діагностичні ознаки розпізнавання, одержаної за алгоритмом (2). Тут темні ділянки графіків позначають робочі (допустимі) області, в яких значення достовірностей перевищують відповідні помилки першого та другого роду.

Аналіз рис. 2 показує, що оптимальні радіуси контейнерів класів розпізнавання дорівнюють: для класу X_1^0 – $d_1^* = 9$, а для класу X_2^0 – $d_2^* = 21$ кодівих одиниць при значеннях нормованого КФЕ

$E_1^* = 1$ та $E_2^* = 0,38$ відповідно. Для класу X_3^0 радіус контейнера обмежений вимірністю простору ознак, тобто дорівнює $d_3^* = 63$, а значення нормованого КФЕ дорівнює $E_3^* = 0,38$.

У режимі екзамену рішення приймалося шляхом визначення максимального значення геометричної функції належності, яка для гіперсферичного класифікатора і реалізацій класу X_m^0 має вигляд:

$$\text{if } d_{m-1}^* < d[x \oplus x_e] < d_m^* \text{ then } \mu_{m,e} = 1; \text{ else } \mu_{m,e} = 0,$$

де $d[x \oplus x_e]$ – кодова відстань між еталонним вектором-реалізацією x та x_e – вектором-реалізацією образу, що розпізнається, d_{m-1}^* – оптимальний радіус контейнеру класу X_{m-1}^0 , d_m^* – класу X_m^0 .

За результатами фізичного моделювання в режимі екзамену СППР підтверджено працездатність та надійність розробленого інформаційного та програмного забезпечення.

Висновки

У рамках ІЕІ-технології розроблено інформаційне, алгоритмічне та програмне забезпечення уні-модального класифікатора з оптимізацією системи

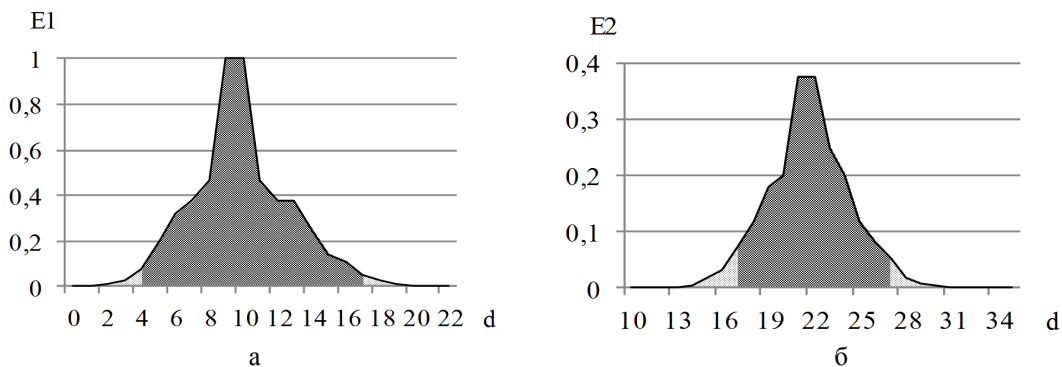


Рис. 2. Графіки залежності нормованого критерію Кульбака від радіусів гіперсферичних контейнерів класів розпізнавання: а – клас X_1^0 ; б – клас X_2^0

контрольних допусків на діагностичні ознаки розпізнавання за паралельним алгоритмом на прикладі діагностування опортуністичних інфекцій у ВІЛ-інфікованих осіб.

З метою підвищення достовірності розпізнавання перспективними є оптимізація інших параметрів функціонування СППР, що впливають на її функціональну ефективність.

Література

1. *Computer-assisted decision support for the diagnosis and treatment of infectious diseases in intensive care units [Text]* / С.А.М. Schurink, Р.Ј.Ф. Lucas, I.М. Hoepelman, М.Ј.М. Bonten // *The Lancet Infectious Diseases*. – 2005. – №5. – С. 305-312.

2. Турбович, И.Т. *Опознание образов. Детерминированно-статистический подход [Текст]* /

И.Т. Турбович, В.Г. Гитис, В.К. Маслов. – М.: Наука, 1971. – 246 с.

3. Довбиш, А.С. *Основи проектування інтелектуальних систем [Текст]: навч. посібник* / А.С. Довбиш. – Суми: Вид-во СумДУ, 2009. – 171 с.

4. Довбиш, А.С. *Паралельно-последовна оптимізація системи контрольних допусків на діагностичні ознаки розпізнавання [Текст]* / А.С. Довбиш, Г.А. Стадник, Г.А. Босенко // *Біоніка інтелекта*. – 2012. – №2(79). – С. 38-42.

5. Довбиш, А.С. *Інформаційно-екстремальний алгоритм унімодального класифікатора для оцінки діаметру монокристалу, що вирощується [Текст]* / А.С. Довбиш, В.В. Москаленко // *Радіоелектронні і комп'ютерні системи*. – 2012. – № 1(53). – С. 114–119.

Надійшла до редакції 02.06.2013, розглянута на редколегії 12.06.2013

Рецензент: д-р техн. наук, проф., завідувач каф. авіаційних приладів та вимірювань М.Д. Кошовий, Національний аерокосмічний університет ім. М.С.Жуковського «ХАІ», Харків.

ИНФОРМАЦИОННО-ЭКСТРЕМАЛЬНЫЙ АЛГОРИТМ ОБУЧЕНИЯ УНИМОДАЛЬНОГО КЛАССИФИКАТОРА ДЛЯ ДИАГНОСТИКИ ОПОРТУНИСТИЧЕСКИХ ИНФЕКЦИЙ У ВИЧ-ИНФИЦИРОВАННЫХ ЛИЦ

А.С. Довбыш, А.А. Стадник, А.И. Поддубная

Предлагается информационно-экстремальный алгоритм анализа и синтеза компьютеризированной системы диагностирования (КСД) инфекционных патологий, построенной на базе унимодального классификатора, который характеризуется единым центром рассеивания векторов-реализаций образов. При этом на этапе обучения системы рассматривается задача построения решающих правил путем построения оптимального в информационном смысле разбиения пространства признаков на классы распознавания с оптимизацией контрольных допусков на диагностические признаки распознавания по параллельному алгоритму. На основании предложенного алгоритма разработано информационное и программное обеспечение системы поддержки принятия решений, которая является основной составляющей КСД, для диагностирования оппортунистических инфекций у ВИЧ-инфицированных лиц.

Ключевые слова: информационно-экстремальный алгоритм, компьютеризированная система диагностирования, унимодальный классификатор, обучение, контрольные допуски, функциональная эффективность, экзамен, оппортунистическая инфекция.

INFORMATION-EXTREME LEARNING ALGORITHM OF UNIMODAL CLASSIFIER FOR DIAGNOSTIC OF OPPORTUNISTIC INFECTIONS IN HIV-INFECTED PEOPLE

A.S. Dovbysh, H.A. Stadnyk, A.I. Piddubna

The information-extreme algorithm of analysis and synthesis of the infectious pathologies computerized diagnostic system (CDS) is proposed. It is based on the unimodal classifier, which is characterized by a single center of patterns vector-realization distribution. Thus the problem of decision rules building is considering by constructing of the feature space optimal partitioning into classes of recognition in the information sense with parallel optimization of the control tolerances for diagnostic recognition attributes. Dataware and software of the decision support system, which is the main component of CDS for diagnosis of opportunistic infections in HIV-infected persons is developed based on the proposed learning algorithm.

Key words: information-extreme algorithm, computer diagnostic system, unimodal classifier, learning, control tolerances, functional efficiency, examination, opportunistic infection.

Довбиш Анатолій Степанович – д-р техн. наук, професор, зав. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Стадник Ганна Анатоліївна – аспірант каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Піддубна Анна Іванівна – аспірант каф. інфекційних хвороб з епідеміологією, Медичний інститут Сумського державного університету, Суми, Україна.