

УДК 681.518:004.93.1'

Г.А. СТАДНИК

Сумський державний університет, Україна

ВИЗНАЧЕННЯ БАЗОВОГО КЛАСУ РОЗПІЗНАВАННЯ ПРИ ДІАГНОСТУВАННІ ІНФЕКЦІЙНИХ ЗАХВОРЮВАНЬ

Розглянуто формування вхідного математичного опису комп'ютеризованої системи діагностування інфекційних патологій у рамках інформаційно-екстремальної інтелектуальної технології, що ґрунтується на максимізації інформаційної спроможності системи в процесі її навчання. При цьому, на етапі навчання системи розглядається задача побудови вирішальних правил шляхом побудови оптимального, в інформаційному розумінні, розбиття простору ознак на класи розпізнавання. Розроблено інформаційне та програмне забезпечення комп'ютеризованої системи діагностування, яке дозволяє на етапі формування вхідного математичного опису автоматично здійснювати вибір базового класу для заданого алфавіту.

Ключові слова: інформаційно-екстремальний алгоритм, комп'ютеризована система діагностування, базовий клас, навчання, функціональна ефективність, гостра кишкова інфекція.

Вступ

Використання сучасного лікувально-діагностичного обладнання дозволяє отримувати великий обсяг діагностичних даних, але одночасно для їх аналізу вимагає від лікаря високої професійної підготовки та наявності практичного досвіду. Тому основною тенденцією розвитку комп'ютеризованих систем діагностування (КСД) є застосування інтелектуальних технологій аналізу даних на основі машинного навчання та розпізнавання образів [1-3]. При цьому у більшості відомих методів не досліджується питання впливу базового класу розпізнавання із заданого алфавіту на функціональну ефективність навчання системи підтримки прийняття рішень (СППР), яка є, як і система збору та візуалізації діагностичних даних обов'язковою складовою КСД. Таким чином, визначення базового класу є самостійним завданням формування вхідного математичного опису, здатної навчатися, СППР.

Одним із перспективних напрямків визначення базового класу та підвищення функціональної ефективності СППР є використання ідей і методів інформаційно-екстремальної інтелектуальної технології (ІЕІ-технології), яка ґрунтується на максимізації інформаційної спроможності системи у процесі її навчання [4-6]. У статті [7] розглядався алгоритм інформаційно-екстремального навчання діагностичної СППР для визначення схеми лікування гострої кишкової інфекції (ГКІ) за результатами лабораторних досліджень, в якій вибір базового класу здійснювався безпосередньо лікарем.

У статті розглядається інформаційно-екстремальний алгоритм навчання СППР з визначенням базового класу.

1. Постановка задачі

Розглянемо формалізовану постановку задачі інформаційного синтезу СППР у рамках ІЕІ-технології. Дано алфавіт класів розпізнавання $\{X_m^0 \mid m = \overline{1, M}\}$, що характеризують M функціональних станів патологічного процесу. Сформовано апріорно класифіковану навчальну матрицю типу «об'єкт-властивість» $\|y_{m,i}^{(j)}\|$, $i = \overline{1, N}$, $j = \overline{1, n}$, де N, n - кількість ознак розпізнавання та реалізацій образу відповідно. Крім того дано вектор параметрів навчання СППР $g = \langle x_m, d_m, \delta \rangle$, де x_m - еталонний вектор-реалізація класу X_m^0 , d_m - радіус контейнера класу X_m^0 , що відновлюється в радіальному базисі простору ознак розпізнавання, і δ - параметр поля контрольних допусків. При цьому задано такі обмеження: x_m - вектор, вершина якого визначає геометричний центр контейнера класу X_m^0 , $d_m \in [0; d(x_m \oplus x_c) - 1]$, де $d(x_m \oplus x_c)$ - кодова відстань центру класу X_m^0 від центру найближчого (сусіднього) до нього класу X_c і параметр поля контрольних допусків $\delta \in [0; \delta_H / 2]$, де δ_H - нормоване (експлуатаційне) поле допусків для відносної шкали вимірювання ознак, яке є областю значень для параметра δ .

Необхідно в процесі навчання оптимізувати координати вектора g , що забезпечують максимальне значення усередненого за алфавітом критерію функціональної ефективності (КФЕ) навчання сис-

теми розпізнавання:

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

де E_m – інформаційний КФЕ навчання СППР розпізнавати реалізації класу X_m^0 ;

$\{k\}$ – впорядкована множина кроків навчання (відновлення контейнерів класів розпізнавання).

На етапі екзамени необхідно з максимальною достовірністю прийняти рішення про належність реалізації образу, що розпізнається, одному із класів алфавіту $\{X_m^0\}$.

2. Алгоритм навчання СППР

Алгоритм навчання за ІЕІ-технологією полягає в реалізації багатоциклическої ітераційної процедури оптимізації структурованих просторово-часових параметрів навчання системи і наближення глобального максимуму усередненого за алфавітом $\{X_m^0\}$ значення КФЕ навчання системи розпізнавання до граничного, що дозволяє побудувати безпомилкові, за навчальною матрицею, вирішальні правила. Як параметри навчання розглянемо геометричні параметри контейнерів класів розпізнавання, побудованих в радіальному просторі ознак розпізнавання та систему контрольних допусків на ознаки розпізнавання.

У статті [7] розглядався алгоритм навчання у рамках ІЕІ-технології, при якому здійснюється оптимізація параметра поля контрольних допусків на ознаки розпізнавання за паралельним алгоритмом

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

де G_δ – допустима область значень параметра δ поля контрольних допусків;

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

G_d – допустима область значень радіуса гіперсферичного контейнера класу розпізнавання.

Вхідними даними для алгоритму паралельної оптимізації системи контрольних допусків на ознаки розпізнавання є масив $\{y_{m,i}^{(j)}\}$ – дійсна, у загальному випадку, багатовимірна навчальна матриця, а як робоча змінна – параметр поля контрольних допусків δ , що дорівнює половині симетричного поля контрольних допусків і значення якого змінюються в інтервалі $[0; \delta_H / 2]$.

Розглянемо алгоритм вибору базового класу:

Обнуляється лічильник порядкових номерів класів розпізнавання $p := 0$.

Формується лічильник порядкових номерів класів розпізнавання: $p := p + 1$.

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

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

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

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

2. Формується множина $\{x_m\}$ еталонних векторів-реалізацій класів розпізнавання за правилом

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

3. Здійснюється розбиття множини $\{x_m\}$ на пари найближчих еталонних векторів.

4. Для кожного класу розпізнавання відновлюється контейнер за схемою:

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

$m := m + 1$;

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

$d := d + 1$;

обчислюються значення інформаційного КФЕ (1);

якщо $E_m \notin G_E$, то виконується пункт 8г, інакше – пункт 8е;

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

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

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

5. Якщо $m \leq M$, то виконується пункт 8б, інакше – пункт 10.

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

7. Якщо $\delta[l] \leq \delta_H / 2$, то виконується пункт 4, інакше – пункт 12.

8. Якщо $\bar{E}[l] \notin G_E$, то виконується пункт 13, інакше – пункт 2.

9. Виконується процедура пошуку глобального максимуму КФЕ $\bar{E}[l]$ в робочій області визначення його функції.

$$10. \bar{E}^*[l] := \text{extrem}_{\{\delta\}} E_m[l].$$

11. Визначається оптимальний параметр поля контрольних допусків на i -ту ознаку розпізнавання: $\delta^* := \arg \bar{E}^*[l]$ і за формулою (2) обчислюються оптимальні контрольні допуски на i -ту ознаку розпізнавання

$$A_{НК_i}^* = y_{1,i} - \delta^* \frac{\delta_{H_i}}{100}; A_{БК_i}^* = y_{1,i} + \delta^* \frac{\delta_{H_i}}{100}.$$

$$12. p := p + 1.$$

13. Якщо $p \leq M$, то виконується пункт 2, інакше – пункт 18.

14. Визначається базовий клас, використання якого забезпечує максимальне усереднене значення КФЕ

$$p^* = \arg \left[\max_{p \in M} \{ \bar{E} \} \right].$$

15. ЗУПИН.

Як критерій оптимізації параметрів навчання у рамках ІЕІ-технології може розглядатися будь-яка статистична інформаційна міра, яка є функціоналом від точнісних характеристик. Наприклад, для двохальтернативних рішень і рівноймовірних гіпотез модифікація критерію Кульбака має вигляд [3]:

$$E_m^{(k)} = \frac{1}{n} \log_2 \left\{ \frac{2n + 10^{-r} - [K_{2,m}^{(k)} + K_{3,m}^{(k)}]}{[K_{2,m}^{(k)} + K_{3,m}^{(k)}] + 10^{-r}} \right\} * [n - (K_{2,m}^{(k)} + K_{3,m}^{(k)})], \quad (4)$$

де $K_{1,m}^{(k)}$ – кількість подій, які означають належність реалізацій образу контейнеру $K_{1,m}^0$, якщо дійсно $\{x_1^{(j)}\} \in X_1^0$;

$K_{2,m}^{(k)}$ – кількість подій, які означають неналежність реалізацій контейнеру $K_{1,m}^0$, якщо дійсно $\{x_1^{(j)}\} \in X_1^0$;

$K_{3,m}^{(k)}$ – кількість подій, які означають належність реалізацій контейнеру $K_{1,m}^0$, якщо вони насправді належать класу X_2^0 ;

$K_{4,m}^{(k)}$ – кількість подій, які означають неналежність реалізацій контейнеру $K_{1,m}^0$, якщо вони насправді належать класу X_2^0 ;

k – крок навчання системи розпізнавання; n_{\min} – мінімальний обсяг репрезентативної навчальної вибірки,

r – число цифр у мантисі критерію $E_m^{(k)}$.

3. Результати фізичного моделювання

Розглянемо застосування алгоритму навчання СППР з вибором базового класу із заданого алфавіту класів розпізнавання для визначення схеми лікування ГКІ, викликані умовно-патогенними мікроорганізмами за результатами лабораторних досліджень. Алфавіт класів розпізнавання складався із трьох класів. При цьому клас X_1^0 характеризує контрольну групу осіб. Клас X_2^0 – групу пацієнтів, для яких необхідне комбіноване лікування з включенням до схеми колоїдного срібла (10 мг/л) і клас X_3^0 – групу пацієнтів, для яких необхідне одночасне призначення пробіотика та колоїдного срібла на тлі базисної терапії. Навчальні матриці класів мали по 40 реалізацій, кожна з яких складалася з 19 ознак розпізнавання. При цьому вектори-реалізації класів подано у вигляді структурованої послідовності ознак розпізнавання, одержаних за результатами лабораторних досліджень мікробіоценозу кишечнику, рівня секреторного IgA, протизапального цитокіну ІЛ 1 β , протизапального цитокіну ІЛ 4, інтегративних показників ендогенної інтоксикації, що вводяться лікарем.

На першій ітерації вищеописаного алгоритму, як базовий, розглядався клас X_1^0 . На рис. 1 показано графік залежності усередненого КФЕ від параметра поля допусків δ , одержаний в процесі паралельної оптимізації системи контрольних допусків на ознаки розпізнавання, яка здійснювалася для всіх ознак одночасно.

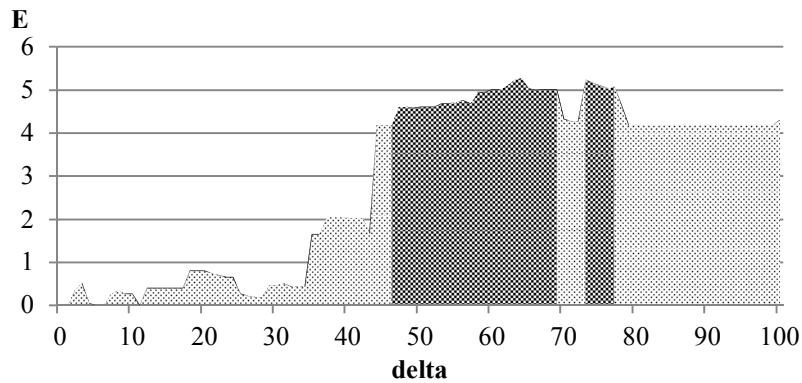


Рис. 1. Графік залежності критерію Кульбака від величини параметра поля контрольних допусків для базового класу X_1^0

На рис. 1 і далі темними ділянками позначено робочі (допустимі) області визначення функції критерію (4), в яких перша та друга достовірності перевищують значення помилок першого та другого роду відповідно. Аналіз рис. 1 показує, що оптимальне значення параметра поля контрольних допусків дорівнює $\delta^* = \pm 64\%$ від усереднених для класу X_1^0 значень ознак розпізнавання при $\bar{E}^* = 5,28$. При цьому оптимальні радіуси контейнерів класів розпізнавання дорівнюють: для класу $X_1^0 - d_1^* = 4$ (тут і далі в кодових одиницях); для класу $X_2^0 - d_2^* = 2$ і для класу $X_3^0 - d_3^* = 3$, а міжцентрові відстані між парами найближчих класів дорівнюють $d(x_1 \oplus x_2) = 8$, $d(x_2 \oplus x_3) = 4$ і $d(x_3 \oplus x_2) = 4$.

На другій ітерації алгоритму, як базовий, приймався клас X_2^0 . На рис. 2 показано графік залежності усередненого КФЕ від параметра δ в процесі

оптимізації контрольних допусків на ознаки розпізнавання.

Аналіз рис. 2 показує, що оптимальне значення параметра поля контрольних допусків на ознаки дорівнює $\delta^* = \pm 97\%$ від усереднених для класу X_1^0 значень ознак розпізнавання при максимальному усередненому значенні КФЕ $\bar{E}^* = 4,84$. При цьому оптимальні радіуси контейнерів класів розпізнавання дорівнюють відповідно $d_1^* = 1$ для класу X_1^0 , $d_2^* = 1$ для класу X_2^0 і $d_3^* = 1$ для класу X_3^0 , а міжцентрові відстані між парами найближчих класів дорівнюють $d(x_1 \oplus x_3) = 2$, $d(x_2 \oplus x_1) = 2$ та $d(x_3 \oplus x_1) = 2$.

На рис. 3 показано графік залежності усередненого КФЕ від параметра δ при виборі класу X_3^0 як претендента на базовий.

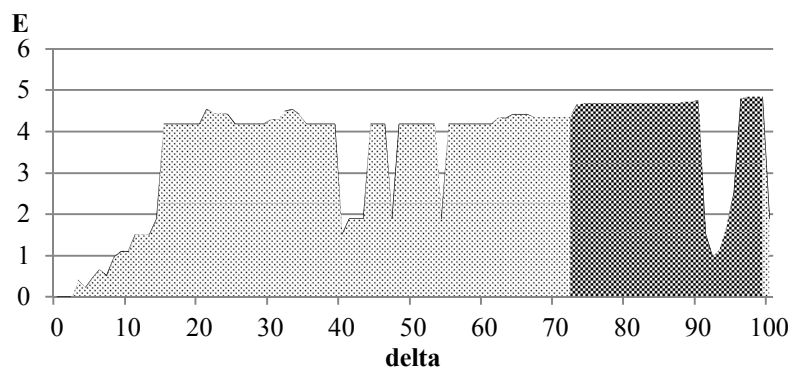


Рис. 2. Графік залежності критерію Кульбака від величини параметра поля контрольних допусків для базового класу X_2^0

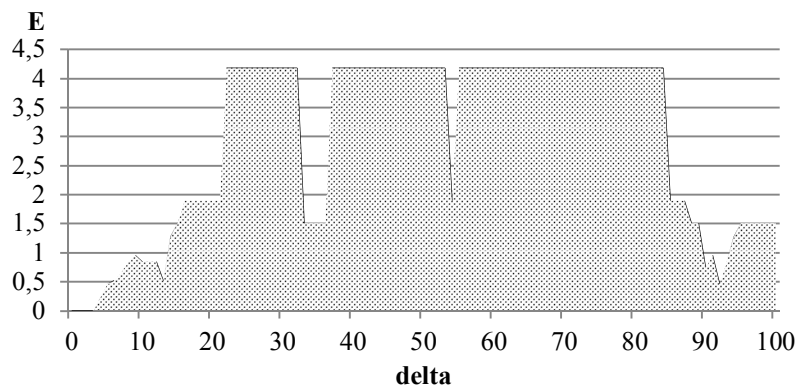


Рис. 3. Графік залежності критерію Кульбака від величини параметра поля контрольних допусків для базового класу X_3^0

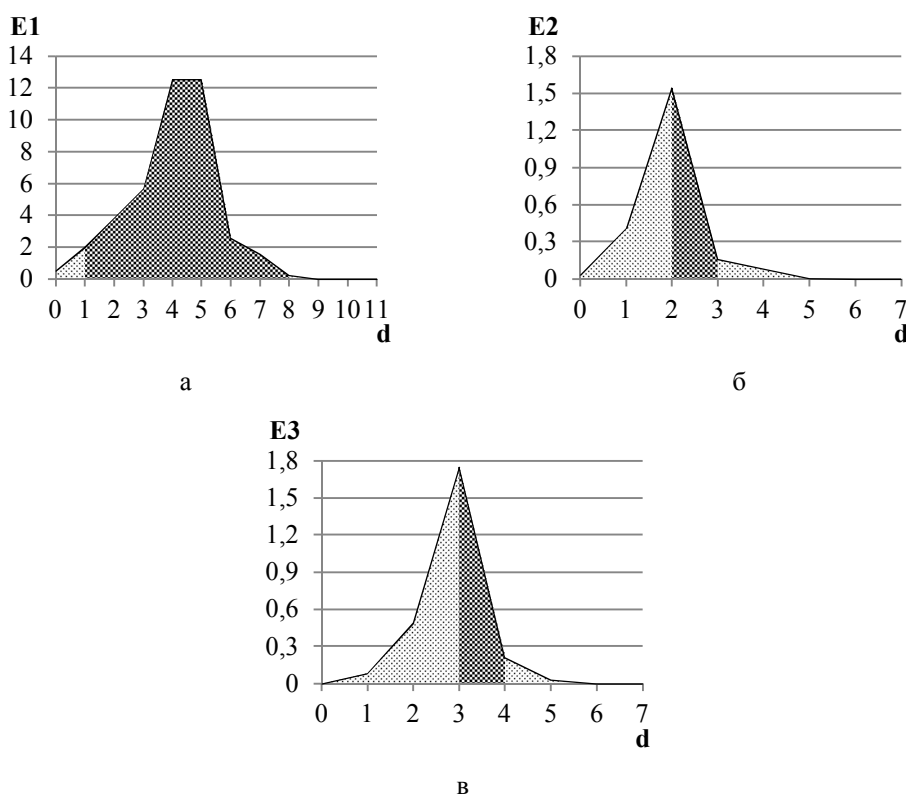


Рис. 4. Графіки залежності критерію Кульбака від радіусів гіперсферичних контейнерів класів розпізнавання: а – клас X_1^0 ; б – клас X_2^0 ; в – клас X_3^0

Аналіз рис. 3 показує, що вибір класу X_3^0 як базового не забезпечує знаходження значення усередненого КФЕ в робочій області визначення його функції.

Таким чином, як базовий клас було обрано клас X_1^0 . На рис. 4 показано графіки залежності КФЕ (4) від радіусів гіперсферичних контейнерів класів розпізнавання, одержаних за умови застосування системи оптимальних контрольних допусків на ознаки розпізнавання при виборі класу X_1^0 як базового.

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

$$\mu_m = 1 - \frac{d(x_m^* \oplus x_e)}{d_m^*},$$

де x_e – вектор-реалізація образу, що розпізнається.

За результатами фізичного моделювання в режимі екзамену СППР підтверджено працездатність

та надійність розробленого інформаційного та програмного забезпечення.

Висновки

Запропоновано інформаційне та програмне забезпечення здатної навчатися СППР для визначення схеми лікування ГКІ за апріорно класифікованою навчальною матрицею у рамках ІЕІ-технології.

Досліджено вплив базового класу для заданого алфавіту класів розпізнавання на функціональну ефективність навчання СППР і визначено оптимальний базовий клас, що дозволило підвищити достовірність діагностування.

З метою підвищення достовірності функціонування СППР перспективними є оптимізація інших параметрів функціонування СППР та перехід на ієрархічну структуру алгоритмів навчання та екзамени, що дозволить збільшити потужність алфавіту класів розпізнавання.

Література

1. Литвин, А.А. Система поддержки принятия решения в прогнозировании и диагностике инфицированного панкреонекроза [Текст] / А.А.Литвин, О.Г.Жариков,

В.А. Ковалев // *Врач и информационные технологии*. – 2012. – №2. – С. 54-63.

2. Усков, А.А. Экспресс-диагностика ОРВИ средствами нечетко-логической экспертной системы [Текст] / А.А. Усков, М.В. Шитилов, В.В. Иванов // *Международный журнал Программные продукты и системы*. – 2011. – № 3. – С. 89-92.

3. A Novel Method for Diagnosing Cirrhosis in Patients with Chronic Hepatitis B: Artificial Neural Network Approach [Text] / Mohammad, Reza Raoufy, Parviz Vahdani, Seyed Moayed Alavian, Sahba Fekri, Parivash Eftekhari, Shahriar Gharibzadeh // *Journal of Medical Systems*. – 2011. – №1(35). – С.121-126

4. Довбиш, А.С. Основы проектирования интеллектуальных систем [Текст]: Навчальний посібник / А.С. Довбиш. – Суми: Видавництво СумДУ, 2009. – 171 с.

5. Шелехов, І.В. Вибір базового класу при розпізнаванні зображень [Текст] / І.В. Шелехов, К.В. Бароіло // *Вісник Сумського державного університету. Серія «Технічні науки»*. – 2010. – №3. – С. 95-102.

6. Довбиш, А.С. Інтелектуальна система діагностування онкопатологій [Текст] / А.С. Довбиш, О.П. Чекалов, С.С. Мартиненко // *Радіоелектронні і комп'ютерні системи*. – 2009. – №3(37). – С. 92-96.

7. Довбиш, А.С. Система підтримки прийняття рішення для визначення схеми лікування гострої кишкової інфекції [Текст] / А.С. Довбиш, Г.А. Стадник, К.С. Полов'яні // *Вісник Сумського державного університету. Серія «Технічні науки»*. – 2012. – №1. – С. 20-26.

Надійшла до редакції 04.11.2012

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

ОПРЕДЕЛЕНИЕ БАЗОВОГО КЛАССА РАСПОЗНАВАНИЯ ПРИ ДИАГНОСТИРОВАНИИ ИНФЕКЦИОННЫХ ЗАБОЛЕВАНИЙ

А.А. Стадник

Рассматривается формирование входящего математического описания компьютеризированной системы диагностирования инфекционных патологий в пределах информационно-экстремальной интеллектуальной технологии, основанной на максимизации информационной способности системы в процессе ее обучения. При этом, на этапе обучения системы рассматривается задача построения решающих правил путем оптимального, в информационном смысле, разбиения пространства признаков на классы распознавания. Разработано информационное и программное обеспечение компьютеризированной системы диагностирования, которое позволяет на этапе формирования входящего математического описания осуществлять выбор базового класса для заданного алфавита.

Ключевые слова: информационно-экстремальный алгоритм, компьютеризированная система диагностирования, обучение, функциональная эффективность, острая кишечная инфекция.

THE SELECTION OF BASE CLASS OF RECOGNITION DURING INFECTION DISEASE DIAGNOSTIC

H.A. Stadnyk

Input mathematical description forming of infectious diseases computerized diagnosing system within the bounds of information-extreme intellectual technology, which is based on maximization of informational capability of recognition system in the process of its learning, is considered. Herewith the issue of decision rules construction by feature space optimal partitioning into classes of recognition in an information sense is considered. Dataware and software of computerized diagnosing system which allows implementing the choice of the base class from the given alphabet at the stage of input mathematical description forming are developed.

Key words: information-extreme algorithm, computer diagnostic system, learning, functional efficiency, acute intestinal infection.

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