

В. О. ДРОФА

ОПТИМІЗАЦІЯ КОНТРОЛЬНИХ ДОПУСКІВ НА ОЗНАКИ РОЗПІЗНАВАННІ ПРИ МАШИННОМУ НАВЧАННІ

Запропоновано в рамках інформаційно-екстремальної інтелектуальної технології аналізу даних алгоритм машинного навчання системи розпізнавання зображень морфології тканин при діагностуванні онкопатологій. При цьому оптимізація контрольних допусків на діагностичні ознаки здійснюється за паралельно-последовательним алгоритмом, що дозволяє підвищити достовірність і оперативність машинного навчання. Для забезпечення інваріантності алгоритму навчання системи розпізнавання до зсуву та повороту оброблення зображень здійснювалося у полярній системі координат. Реалізація алгоритму розглянуто на прикладі розпізнавання цифрових морфологічних зображень тканин, отриманих методом біопсії.

Ключові слова: система розпізнавання, оптимізація, контрольні допуски, машинне навчання, зображення, інформаційний критерій, функціональна ефективність.

Предложено в рамках информационно-экстремальной интеллектуальной технологии анализа данных алгоритм машинного обучения системы распознавания изображений морфологии тканей при диагностировании онкопатологий. При этом оптимизация контрольных допусков на диагностические признаки осуществляется параллельно-последовательным алгоритмом, позволяющим повысить достоверность и оперативность машинного обучения. Для обеспечения инвариантности алгоритма обучения системы распознавания к смещению и поворота обработки изображений осуществлялось в полярной системе координат. Реализация алгоритма рассмотрена на примере распознавания цифровых морфологических изображений тканей, полученных методом биопсии.

Ключевые слова: система распознавания, оптимизация, контрольные допуски, машинное обучение, изображения, информационный критерий, функциональная эффективность.

The article present method under extreme intellectual information technologies of data analysis algorithm machine learning image recognition system tissue morphology in the diagnosis cancer. This optimization control tolerances diagnostic signs carried by parallel-sequential algorithm, which improves the reliability and efficiency of machine learning to provide training algorithm invariance recognition system to shift and rotate images processing was carried out in a polar coordinate system. As functional efficiency criterion used Shannon entropy measure. Application of such algorithm allows to consider the quasioptimal control tolerances received in the course of parallel optimization, as start for serial algorithm. It guarantees a finding in the course of optimization values of entropy criterion of the functional learning efficiency in working (admissible) area determination of its function. Implementing the algorithm considers the example of morphological recognition of digital images of tissue obtained by biopsy.

Keywords: recognition system, optimization, control tolerances, machine learning, image, information criterion, functional efficiency.

Вступ. Основною тенденцією розвитку комп'ютеризованих систем діагностування є застосування інтелектуальних технологій аналізу даних на основі машинного навчання та розпізнавання образів [1, 2]. При цьому складність розпізнавання цифрових морфологічних зображень тканин пацієнтів, отриманих за методом біопсії, викликана такими факторами, як багатовимірність словника діагностичних ознак і алфавіту класів розпізнавання, довільні початкові умови формування зображень і суттєвий перетин класів розпізнавання, що обумовлює априорно нечітке розбиття простору діагностичних ознак на класи розпізнавання. Для підвищення функціональної ефективності систем розпізнавання зображень медичних і біологічних об'єктів перспективним є використання ідей і методів інформаційно-екстремальної інтелектуальної технології (ІЕІ-технології) [3, 4], що базується на максимізації інформаційної спроможності здатної навчатися системи розпізнавання, яка є основною складовою інтелектуальної комп'ютеризованої системи діагностування. У праці [5] в рамках ІЕІ-технології розглядався так званий базовий алгоритм машинного навчання системи розпізнавання морфологічних зображень тканин пацієнтів для розпізнавання раку молочної залози, але отримані результати не дозволили досягнути високої функціональної ефективності машинного навчання через неоптимальні контрольні допуски на діагностичні ознаки.

У статті розглядається інформаційно-екстре-

мальний алгоритм машинного навчання системи розпізнавання з паралельно-последовательною оптимізацією контрольних допусків на ознаки розпізнавання.

Постановка задачі. Розглянемо здатну навчатися систему розпізнавання морфологічних зображень тканин пацієнтів, отриманих методом біопсії. Нехай дано у загальному випадку нечіткий алфавіт класів розпізнавання $\{X_m^o, m = \overline{1, M}\}$, які характеризують як доброякісні, так і злякісні онкопатології. Для заданого алфавіту класів розпізнавання сформовано вхідну навчальну матрицю яскравості пікселів рецепторного поля зображень типу «об'єкт-властивість»

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

де N, n – кількість діагностичних ознак і кількість векторів-реалізацій образу відповідно. Для системи розпізнавання відомий структурований вектор параметрів навчання

$$g_m = \langle g_{m,1}, \dots, g_{m,q}, \dots, g_{m,Q} \rangle \quad (1)$$

із відповідними обмеженнями.

В режимі навчання необхідно знайти оптимальні (тут і далі в інформаційному розумінні) координати вектора параметрів навчання (1) шляхом пошуку глобального максимуму інформаційного критерію функціональної ефективності (КФЕ) в робочій області визначення його функції:

$$\bar{E}^* = \frac{1}{M} \sum_{m=1}^M E_m^* \quad (2)$$

де E_m^* – максимальне значення КФЕ навчання системи діагностування розпізнавати реалізації класу X_m^0 .

Опис алгоритму. Оскільки контрольні допуски на значення ознак розпізнавання прямо впливають на геометричні параметри контейнерів класів розпізнавання, а таким чином і на асимптотичні точнісні характеристики рішень, то питання оптимізації системи контрольних допусків (СКД) у методах ІЕІ-технології набуває важливого значення при розробленні інформаційного забезпечення системи, що навчається.

Розглянемо алгоритм інформаційно-екстремального машинного навчання системи розпізнавання зображень з оптимізацією контрольних допусків на ознаки розпізнавання. На рис. 1 показано симетричне (двобічне) поле допусків на значення i -ї ознаки розпізнавання y_i .

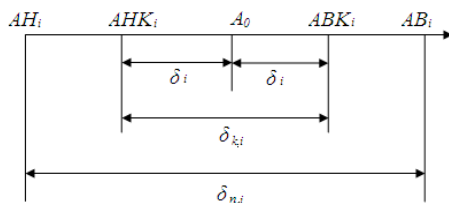


Рис. 1 – Симетричне поле допусків на ознаку розпізнавання

На рис. 1. прийнято такі позначення:

A_0 – номінальне значення ознаки y_i ;

AN_i, AB_i – нижній і верхній нормовані допуски відповідно;

ANK_i, ABK_i – нижній і верхній контрольні допуски відповідно;

$\delta_{n,i}$ – нормоване поле допусків;

$\delta_{k,i}$ – контрольне поле допусків;

δ_i – параметр поля симетричний центру допусків, який дорівнює половині його довжини.

Категорійна модель відображень множин у процесі машинного навчання з оптимізацією контрольних допусків на ознаки розпізнавання показана на рис. 2.

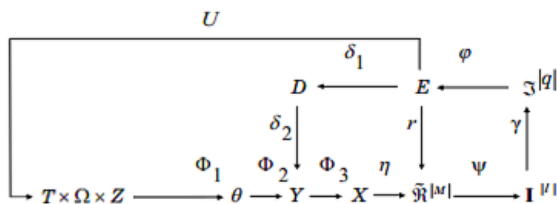


Рис.2 – Структурна діаграма процесу машинного навчання

На рис.2 прийнято такі позначення:

Y – множина сигналів, які знімаються з виходу блоку первинної обробки інформації;

$\bar{\mathfrak{R}}^{|M|}$ – нечітке розбиття;

$I^{|I|}$ – множина можливих рішень,

E – множина значень інформаційного КФЕ;

$\mathfrak{Z}^{|q|}$ – множина точнісних характеристик;

θ – спектр яскравості зображення;

Φ_1 – оператор обробки зображень в полярних координатах;

Φ_2 – оператор формування множини Y ;

Φ_3 – оператор формування навчальних матриць.

Контур, в якому знаходиться впорядкована термножина D оптимізує систему контрольних допусків на ознаки розпізнавання в процесі цілеспрямованого пошуку глобального максимуму критерію (2). Оператор U регламентує процес машинного навчання.

Оптимізацію контрольних допусків на ознаки розпізнавання будемо здійснювати за паралельно-послідовним алгоритмом.

Перевагою паралельного алгоритму (3) оптимізації СКД є висока оперативність реалізації алгоритму, але він не дозволяє одержати точне значення глобального максимуму КФЕ в робочій області визначення його функції. Тому екстремальні значення параметрів навчання системи, одержані в процесі їх оптимізації за паралельним алгоритмом, є квазіоптимальними.

Алгоритм послідовної (4) оптимізації СКД дозволяє обчислювати точні значення глобального максимуму КФЕ в робочій області, але характеризується низькою оперативністю.

З метою поєднання переваг цих алгоритмів оптимізацію СКД на ознаки розпізнавання доцільно здійснювати на практиці за паралельно-послідовним алгоритмом. При цьому реалізація паралельного алгоритму дозволяє визначити стартові контрольні допуски, які є вхідними для алгоритму послідовної оптимізації. Це дозволяє підвищити оперативність послідовного алгоритму, оскільки стартові квазіоптимальні контрольні допуски вже знаходяться в робочій області визначення функції КФЕ.

Алгоритм оптимізації контрольних допусків, як і інших параметрів функціонування системи розпізнавання, у рамках ІЕІ-технології полягає у наближенні глобального максимуму інформаційного критерію оптимізації, що обчислюється в робочій області визначення функції, до граничного найбільшого його значення.

Алгоритм паралельної оптимізації контрольних допусків на ознаки розпізнавання здійснюється за ітераційною процедурою пошуку глобального значення інформаційного КФЕ в робочій (допустимій) області визначення його функції:

$$\{\delta_{k,i}^* \mid i = \overline{1, N}\} = \arg \max_{G_\delta} \max_{G_E \cap G_d} \bar{E}, \quad (3)$$

де $\delta_{k,i}^*$ – оптимальне значення параметра поля контрольних допусків;

G_δ, G_E, G_d – області допустимих значень контрольних допусків на ознаки розпізнавання, інформаційного КФЕ і радіусів гіперсферичних контейнерів класів розпізнавання, що відновлюються у процесі навчання в радіальному базисі простору ознак відповідно.

Алгоритм послідовної оптимізації поля контрольних допусків на ознаки розпізнавання має вигляд:

$$\{\delta_{K,i}^*\} = \langle \arg\{\max_{G_{\delta_i}} \{\max_{G_E} [\otimes_{l=1}^L \max_{G_{d_l}} E_1^{(l)}]\}\} \rangle, i = \overline{1, N}, (4)$$

де $G_{\delta_i}, G_E, G_{d_l}$ – області допустимих значень поля контрольних допусків для i -ї ознаки, критерію оптимізації і радіуса d_l контейнера класу X_1^0 відповідно; \otimes – символ операції повторення.

Функціональна ефективність є важливою характеристикою системи діагностування, що навчається. Функціональна ефективність, як складова загальної ефективності, визначає ступінь відповідності функціонування системи за її робочим алгоритмом виконання поставленої перед нею задачі, згідно з критерієм мети. Важливою складовою критерію мети є інформаційний критерій функціональної ефективності (КФЕ) навчання системи, що є функціоналом від точнісних характеристик рішень, які приймаються системою.

Результати. Розглянемо результати реалізації запропонованого алгоритму на прикладі навчання СППР на етапі діагностування морфологічних зображень тканин пацієнтів, отриманих методом біопсії.

З метою забезпечення інваріантності алгоритму навчання КСД до зсуву та повороту зображень їх оброблення здійснювалося в полярній системі координат, в якій значення яскравості пікселів рецепторного поля зображення зчитуються за правилами

$$\begin{aligned} x &= r \cos \varphi; \\ y &= r \sin \varphi, \end{aligned} (5)$$

де r, φ – радіус і кут кола зчитування відповідно.

Формування навчальної матриці здійснювалося шляхом оброблення зображень в полярних координатах за формулою

$$\Theta_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \theta_i, (6)$$

де Θ_j – усереднене значення яскравості пікселів, що належать колу зчитування j -го радіусу, $j = \overline{0, R}$;

θ_i – значення яскравості кольорової складової в i -му пікселі;

N_j – загальна кількість пікселів в j -му колі зчитування.

Усереднені значення яскравості в кожному колі зчитування розглядаються як координати структурованого вектора – реалізації зображення. Оскільки зображення морфології тканин є кольоровими, то їх оброблення здійснювалося в форматі RGB. При цьому навчальна матриця складалася відповідно із трьох RGB-складових.

Алфавіт класів розпізнавання складається з чотирьох зображень, що характеризують онкологічні захворювання.

У табл. 1 показано приклади зображень морфології тканин, отримані за методом біопсії.

Таблиця 1 – Алфавіт класів розпізнавання

№ класу	Позначення	Назва	Зображення
1	X_1^0	Інтраканалікулярна фіброаденома	
2	X_2^0	Філоїдна фіброаденома псевдосаркоматозна	
3	X_3^0	Периканалікулярна непроліферативна фіброаденома	
4	X_4^0	Мазоплазія	

Для оцінки функціональної ефективності СППР використовувався інформаційний критерій Шеннона

$$\begin{aligned} E &= 1 + \frac{1}{2} \left(\frac{D_1}{D_1 + \beta} \log \frac{D_1}{D_1 + \beta} + \frac{\beta}{D_1 + \beta} \log \frac{\beta}{D_1 + \beta} + \right. \\ &\quad \left. + \frac{1 - D_1}{2 - D_1 - \beta} \log \frac{1 - D_1}{2 - D_1 - \beta} \right), \end{aligned} (7)$$

де D_1, β – перша достовірність та помилка другого роду відповідно.

На рис.3. показано графік залежності ентропійного критерію (7) від параметра δ поля контрольних допусків на ознаки розпізнавання, отриманий при паралельній їх оптимізації.

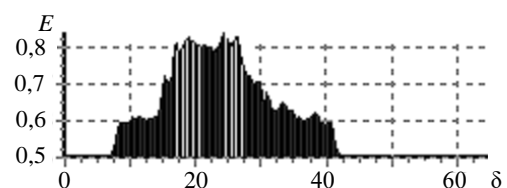


Рис. 3 – Графік залежності КФЕ від параметра поля контрольних допусків на ознаки розпізнавання

Аналіз рис. 3 показує, що оптимальний параметр поля контрольних допусків дорівнює $\delta_{K,i}^* = \pm 24$, а максимальне значення інформаційного критерію (7) – $\overline{E}^* = 0,86$.

На рис. 4 показано графіки залежності критерія функціональної ефективності (КФЕ) від контейнера класу за паралельним алгоритмом. На графіку штриховані ділянки позначають робочі (допустимі) області визначення функції інформаційного критерію, в яких одночасно виконуються умови: перша достовірність і помилка другого роду приймають значення $D_1 > 0,5$ і $\beta < 0,5$ відповідно.

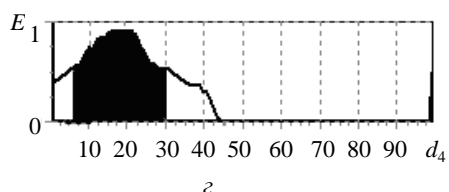
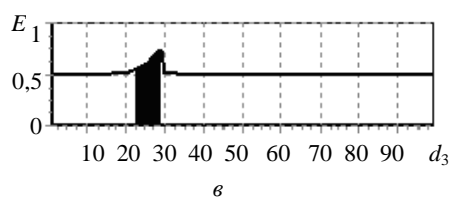
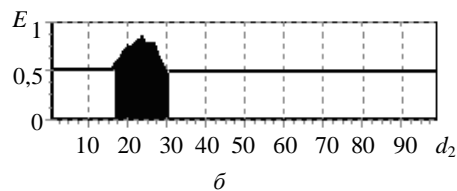
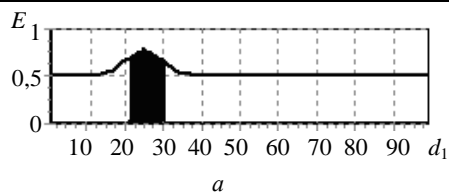


Рис. 4 – Графіки залежності КФЕ від радіусів контейнерів класів розпізнавання: а – X_1^0 , б – X_2^0 , в – X_3^0 , з – X_4^0

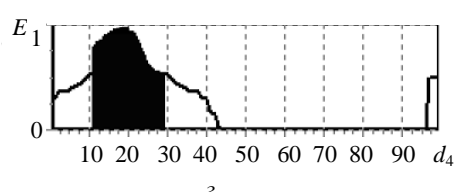
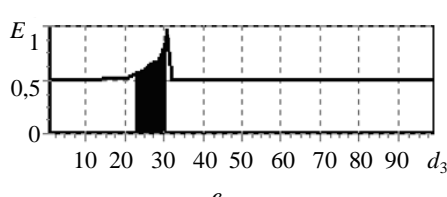
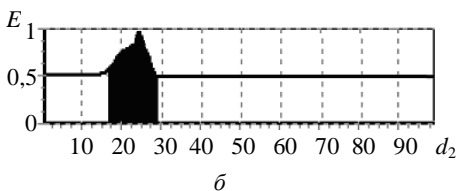
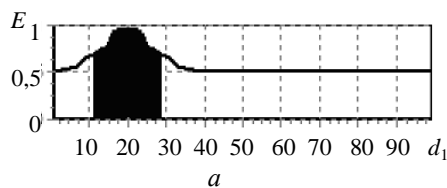


Рис. 5 – Графіки залежності КФЕ від радіусів контейнерів класів розпізнавання: а – X_1^0 , б – X_2^0 , в – X_3^0 , з – X_4^0

В табл.2 подано оптимальні параметри для класів розпізнавання за паралельним алгоритмом.

Таблиця 2 – Оптимальні параметри навчання системи розпізнавання

Клас	КФЕ	Радіус	D_1	β
X_1^0	0,76	24	1,00	0,03
X_2^0	0,83	24	0,93	0,00
X_3^0	0,74	29	0,87	0,00
X_4^0	0,88	22	0,97	0,00

На рис. 5 показано графіки залежності критерія функціональної ефективності (КФЕ) від контейнера класу за паралельно-последовним алгоритмом.

На рис. 6 показано графік залежності КФЕ від кількості ітерацій алгоритму навчання СППР та отримано усереднене значення КФЕ навчання $\bar{E}^* = 0,98$.

В табл.3 подано оптимальні параметри для класів розпізнавання за паралельно-последовним алгоритмом.

Аналіз отриманих результатів за паралельно-последовним алгоритмом свідчить про те, що у порівнянні з паралельним алгоритмом, він дає кращі результати параметрів навчання та дозволяє підвищити оперативність оптимізації системи контрольних допусків на діагностичні ознаки при забезпеченні максимальної асимптотичної достовірності навчання СППР.



Рис. 6 – Графік залежності КФЕ від кількості ітерацій алгоритму навчання СППР

Таблиця 3 – Оптимальні параметри навчання системи розпізнавання

Клас	КФЕ	Радіус	D_1	β
X_1^0	0,98	23	1,00	0,03
X_2^0	0,98	25	0,97	0,00
X_3^0	0,98	31	0,97	0,00
X_4^0	0,98	22	0,97	0,00

Висновки. В роботі запропоновано категорійну модель і на її основі реалізовано алгоритм інформаційно-екстремального машинного навчання системи розпізнавання нестационарних за яскравістю зображень із оптимізацією системи контрольних допусків на ознаки розпізнавання. При цьому формування вхідної навчальної матриці здійснювалося шляхом оброблення зображень в полярній системі координат,

що дозволило забезпечити інваріантність алгоритму навчання до їх зсуву і повороту

Отримані в процесі фізичного моделювання результати показали, що реалізація алгоритму інформаційно-екстремального машинного навчання системи діагностування онкопатологій за зображеннями морфології тканин з паралельно-послідовною оптимізацією ознак розпізнавання дозволяє підвищити достовірність діагностичних рішень у порівнянні з паралельною оптимізацією.

Список літератури

1. *Литвин А. А.* Система поддержки принятия решения в прогнозировании и диагностике инфицированного панкреонекроза / *А. А. Литвин, О. Г. Жариков, В. А. Ковалев* // Врач и информационные технологии. – 2012. – № 2. – С. 54–63.
2. *Novel A.* Method for Diagnosing Carcinoma in Patients with Chronic Hepatitis B: Artificial Neural Network Approach / *Mohammad, Reza Raoufy, Parviz Vahdani, Seyed Moayed Alavian, Sahba Fekri, Parivash Eftekhari, Shahriar Gharibzaden* // *Journal of Medical System.* – 2011. – № 1 (35). – P. 121–126.
3. *Краснояковский А. С.* Информационный синтез интеллектуальных систем управления: Подход, что основывается на методі функціонально-статистичних випробувань / *А. С. Краснояковский* // Суми: Вид-во СумДУ, 2004. – 261 с.
4. *Dovbysh A. S.* Information-extreme algorithm for recognizing current distribution maps in magnetocardiography / *A. S. Dovbysh,*

- S. S. Martynenko, A. S. Kovalenko, N. N. Bydnyk* // *Journal of automation and information sciences.* – 2011 – 43 (2) – P. 63 – 70.
5. *Довбиш А. С.* Основы проектирования интеллектуальных систем: навч. посібник / *А. С. Довбиш.* Суми : Вид-во СумДУ, 2009 – С.171.

References (transliterated)

1. *Lytvyn A. A., Zharykov O. G., Kovalev V. A.* Systema podderzhky prinyatiya resheniya v prognozyrovaniy u dyagnostyke ynfycirovannogo pankreonekroza [Decision support system in forecasting and diagnosis of infected pancreatic necrosis]. *Vrach y ynfomacyonnye tehnologi.* 2012, no. 2, pp. 54–63.
2. *Novel A.* Method for Diagnosing Carcinoma in Patients with Chronic Hepatitis B: Artificial Neural Network Approach. *Journal of Medical System.* 2011, no. 1 (35), pp. 121–126.
3. *Krasnopoyasovskiy A. S.* *Informatsionnyy sintez intelektualnykh system keruvannya: Pidxid, shho gruntuyetsya na metodi funktsionalno-statystychnykh vyprobuvan* [Information synthesis of intelligent control systems: An approach based on the method of functional and statistical tests]. Sumy, SumDU Publ., 2004. 261 p.
4. *Dovbysh A. S., Martynenko S. S., Kovalenko A. S., Bydnyk N. N.* Information-extreme algorithm for recognizing current distribution maps in magnetocardiography. *Journal of automation and information sciences.* 2011, no. 43 (2), pp. 63–70.
5. *Dovbysh A.S.* *Osnovy proektuvannya intelektualnykh system: navch. posibnyk* [Fundamentals of Intelligent Systems]. Sumy, SumDU Publ., 2009. 171 p.

Надійшла (received) 28.11.2016

Бібліографічні описи / Библиографические описания / Bibliographic descriptions

Оптимізація контрольних допусків на ознаки розпізнаванні при машинному навчанні / В. О. Дрофа // Вісник НТУ «ХПІ». Серія: Системний аналіз, управління та інформаційні технології. – X. : НТУ «ХПІ», 2016. – № 45 (1217). – С. 29–33. – Бібліогр.: 5 назв. – ISSN 2079-0023.

Оптимизация контрольных допусков на признаки распознавания при машинном обучении / В. А. Дрофа // Вестник НТУ «ХПИ». Серія: Системный анализ, управление и информационные технологии. – X.: НТУ «ХПИ», 2016. – № 45 (1217). – С. 29–33. – Библиогр. : 5 названий. – ISSN 2079 0023.

Optimization control tolerances for signs of recognition in machine learning / V. A. Drofa // Bulletin of NTU "KhPI". Series: System analysis, control and information technology. – Kharkov : NTU "KhPI", 2016. – No. 45 (1217). – P. 29–33. – Bibliogr.: 5. – ISSN 2079-0023.

Відомості про авторів / Сведения об авторах / About the Authors

Дрофа Вікторія Олександрівна – аспірант кафедри комп'ютерних наук, Сумський державний університет, м. Суми; тел.: (096) 653-55-69; e-mail: brviktoria@yahoo.com.

Дрофа Виктория Александровна – аспірант кафедри комп'ютерних наук, Сумской государственной университет, г. Сумы, тел.: (096)) 653-55-69; e-mail: brviktoria@yahoo.com.

Drofa Victoria A. – graduate student Computer Science Department, Sumy State University, Sumy; tel. : (096) 653-55-69; e mail: brviktoria@yahoo.com.