

МЕТОДОЛОГІЯ СИНТЕЗУ КОЛЕКТИВУ РАДІАЛЬНО-БАЗИСНИХ МЕРЕЖ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧ КЛАСИФІКАЦІЇ ЗА ДОПОМОГОЮ АЛГОРИТМУ КЛОНАЛЬНОГО ВІДБОРУ

У роботі запропонована методологія синтезу колективу радіально-базисних нейронних мереж для розв'язання задач класифікації. На прикладі тестових еталонних задач показана висока ефективність запропонованого підходу.

Ключові слова: *радіально-базисні нейронні мережі, класифікація, алгоритм клонального підходу.*

В работе предложена методология синтеза коллектива радиально базисных нейронных сетей для решения задач классификации. На примере тестовых эталонных задач показана высокая эффективность предложенного подхода.

Ключевые слова: *радиально базисные нейронные сети, классификация, алгоритм клонального подхода.*

In work the methodology of synthesis of collective of RBF- networks for the decision of problems of classification is offered. On an example testovux reference problems high efficiency of the offered approach is shown.

Key words: *radially-base neuron networks, classification, algorithm to clonal approach.*

ВСТУП

Складність задач, що розв'язуються, на сьогоднішній день така, що, застосовуючи тільки один певний алгоритм, одиночну нейронну мережу, часто не вдається домогтися бажаної якості розв'язання. Тому все більше поширення одержують так звані колективи (модулі) нейронних мереж, що представляють собою довільний набір різних алгоритмів обробки даних, включаючи алгоритми штучних нейронних мереж, об'єднаних для розв'язання складної багато-параметричної задачі. Застосування модульних нейронних мереж може бути обґрунтоване з декількох точок зору. З біологічної точки зору існування модулів виявлене на всіх рівнях біологічної ієрархії. Доведено, що мозок являє собою модульну структуру та складається з декількох анатомічно і функціонально різних областей [1]. Дослідження нейроанатомії головного мозку показують, що архітектура головного мозку використовує модульність як ключовий принцип на всіх рівнях його побудови і функціонування. З погляду психології, багато систем, що навчаються, мають функціональні ознаки модульності. При обробці інформації людиною складна задача часто поділяється на менші по розміру частини або на більш легкі підзадачі. З обчислювальної точки зору при навчанні одиночних нейронних мереж для розв'язання комплексних задач виникає велика кількість проблем. Спроби розв'язання складних класифікаційних задач із вкладеними класами або навчання однієї мережі розв'язанню різних незалежних задач часто виявляються невдалими [2]. Застосування в цьому випадку колективів нейронних мереж, що дозволяють розділити процес навчання між нейронними мережами, виявляється особливо корисним. Зокрема, колективи нейронних мереж дозволяють розбивати складні задачі розпізнавання на прості підзадачі, які розв'язуються окремими підмережами, і конструювати розв'язання проблеми в цілому як сукупність

розв'язаних підзадач. Також у багатьох випадках використання модульності та поділ на менші підзадачі дозволяє поліпшити узагальнюючу здатність мережі, прискорити процес навчання, зменшити складність і розміри мереж, що використовуються.

1. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ

Якість розв'язання конкретної поставленої задачі (емуляції, прогнозування, розпізнавання образів, зворотного моделювання, керування і т.п.) може бути істотно підвищена за допомогою колективів (ансамблів) нейромереж [3, 4, 5, 6-10], в яких ті самі дані паралельно обробляються кількома ШНМ, вихідні сигнали яких далі деяким чином комбінуються в об'єднану оцінку, що перевершує по якості результати, одержувані за допомогою локальних мереж. Використання колективу мереж часто дозволяє зробити роботу мережі й отримане з її допомогою відображення «вхід-вихід» більш зрозумілими.

Останнім часом в галузі модульних нейронних мереж отримана велика кількість багатообіцяючих результатів. Були розроблені різні архітектури модульних мереж, способи їхньої оптимізації [12] і навчання [13]. Була розроблена архітектура локальних експертів [14], визначені ансамблеві та власне модульні нейронні мережі [15]. Досліджено різні методи об'єднання результатів окремих модулів [16]. Модульні мережі використовувалися для рішення задач класифікації та інтерполяції, у тому числі таких, як ідентифікація голосу [17], моделювання керування роботом [9], класифікації зображень [10] і типів ракових клітин [18] і т.д.

На практиці найбільше поширення одержали два підходи до об'єднання мереж в ансамблі: модульний [22] і заснований на зваженому усередненні [23]. І хоча змістовно вони досить відрізняються друг від друга, їх поєднує те, що обидва підходи використовують лінійну комбінацію вихідних сигналів своїх членів у тій або іншій формі. Модульний підхід має досить евристичний характер на відміну від більш математично строгого зваженого усереднення, хоча і тут залишається елемент суб'єктивізму, пов'язаний з вибором членів ансамблю. Ця задача зазвичай розв'язується за допомогою тих або інших евристик, хоча і є більш-менш строгі результати, засновані на генетичному програмуванні [16] або поступовому нарощуванні складності мереж-членів ансамблю [1]. Використання нейронної метамережі дозволяє підвищити якість обробки інформації на основі ансамблю ШНМ і визначити мережі-переможці, щонайкраще пристосовані до розв'язання поставленої задачі та які роблять найбільший внесок у її розв'язання.

Існування великої кількості різноманітних обчислювальних парадигм неминуче приводить до ідеї про можливе об'єднання підходів з метою створення гібридів, що сполучають у собі кращі боки парадигм, що беруть участь в об'єднанні. Поява штучних імунних систем обумовила появу подальших розробок у напрямку їхньої гібридизації з іншими обчислювальними системами. Наприклад, зі штучними нейронними мережами, еволюційними алгоритмами, нечіткими системами і т.п. Ідея гібридизації ґрунтується на тому факті, що більшість парадигм містять елементи, інформація про які для розв'язання конкретної задачі вимагає попереднього визначення, а це, у свою чергу, є непростю задачею. Наприклад, в нейронній мережі необхідне визначення її структури, тобто кількості зв'язків, кількості нейронів у шарі та кількості шарів; еволюційні алгоритми повинні мати попередню інформацію про кількість хромосом у популяції, а однією з основних проблем нечітких систем є визначення числа термів лінгвістичної змінної.

Особливе місце займає проблема адаптивного настроювання параметрів, таких як, наприклад, імовірності застосування еволюційних операторів, форми і параметрів функцій приналежності і т.д. Дані задачі по суті можна віднести до класу комбінаторних задач і від їхнього ефективного розв'язання залежить ефективність обчислювальної системи в цілому.

У ряді праць [24-26] нами запропоновано використовувати штучні імунні системи для синтезу нейронних мереж, призначених для розв'язання задач прогнозування. Штучні імунні системи являють собою евристичні алгоритми, однією з переваг яких є здатність до швидкого

розв'язання подібних задач. Тому цілком логічною представляється спроба об'єднання підходів з метою одержання більш якісних результатів.

2. ПОСТАНОВКА ЗАДАЧІ

У даній роботі ставиться задача розробки методології синтезу радіально-базисних нейронних мереж у вигляді бінарних класифікаторів, які синтезуються і навчаються за допомогою імунного алгоритму клонального відбору, досліджується якість роботи даних класифікаторів і досліджується вплив загального рівня мутації на якість роботи класифікатора.

3. РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

3.1 Нейронні мережі радіального базису і проблема настроювання параметрів

Нейронні мережі радіального базису (РБФ-мережі) [27, 28, 29, 30] є потужним засобом апроксимації багатомірних нелінійних функцій. Вони мають досить просту архітектуру і мають високу швидкість навчання. Блок-діаграма РБФ-мережі представлена на малюнку 1. РБФ-мережа складається із вхідного, єдиного прихованого (радіального базисного) і лінійного (вихідного) шарів. Вхідний шар складається із сенсорів, що з'єднують мережу із зовнішнім середовищем. Нейрони прихованого шару діють за принципом центрування на елементах навчальної вибірки. В якості центрів виступає вагова матриця (W^r). У блоці (dist) відбувається обчислення Евклідової відстані між вхідним вектором (X) і відповідним центром. Навколо кожного центра існує область, називана радіусом. Радіус (чутливість мережі) коректується за допомогою вектора коефіцієнтів згладжування: $(\sigma_1, \dots, \sigma_m)$. Функція перетворення (звичай

Гаусова – $f(x) = e^{-\frac{(x-c)^2}{2\sigma^2}}$), що змінюється в інтервалі від 0 до 1, визначає вихід прихованого шару. Вихідний шар містить звичайні лінійні або сигмоїдні нейрони і шляхом настроювання своїх ваг (W^l) визначає вихід мережі.

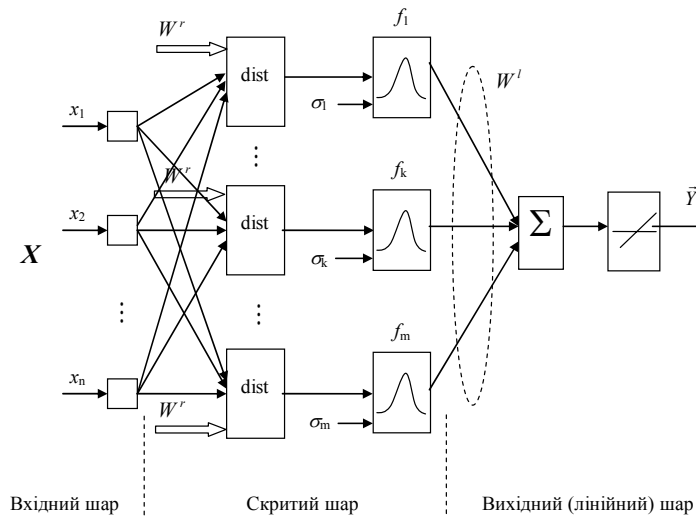


Рис. 1. Узагальнена архітектура РБФ-мережі

Поведінка РБФ-мережі багато в чому залежить від кількості та положення радіальних базисних функцій прихованого шару. Дійсно, для будь-якого дійсного n -мірного вхідного вектора $x = (x_1, x_2, \dots, x_n)$, де $x \in X \subset \mathfrak{R}_n$, вихід мережі буде визначатися в такий спосіб:

$$y_i = \sum_{k=1}^m w_{ik}^l f_k(\text{dist}(x, w_k^r), \sigma_k), \quad (1)$$

де $w_{ik}^l \in W^l$, $i = \overline{1, p}$ – ваги лінійного шару; $w_k^r \in W^r$ – центри радіально-базисних функцій. Якщо в якості базисної функції використовується функція Гауса, то

$$f_k(x) = -\frac{\text{dist}(x, w_k^r)^2}{2\sigma_k^2}, \quad k = \overline{1, m}. \quad (2)$$

У контексті задачі класифікації настроювання мережі полягає в пошуку функції $y: \mathfrak{R}_n \rightarrow \mathfrak{R}$, що задовольняє рівнянню (1) при $p = 1$. Нехай ϵ вибірка, що складається з S навчальних точок даних: $X_1, \dots, X_S, X_i \in \mathfrak{R}_n$. Якщо відомі вихідні значення для кожної із цих точок $d_1, \dots, d_S, d_i \in \mathfrak{R}$, то кожна базисна функція може бути центрована на одній із точок X_i . Отже, у граничному випадку, число центрів i , відповідно, нейронів прихованого шару, буде рівним числу точок даних навчальної вибірки: $m = S$.

У роботі [31] показано, що в цьому випадку виникає як мінімум дві проблеми. По-перше, низька здатність до узагальнення, обумовлена наявністю надмірно великої кількості нейронів прихованого шару, негативно позначається на апроксимації нових (які не брали участь у навчанні) даних, а по-друге, великий розмір навчальної вибірки неминуче викликає проблеми обчислювального характеру. Для подолання зазначених труднощів складність мережі повинна бути зменшена скороченням кількості базисних функцій, що у свою чергу ставить нову задачу, яка полягає в їхньому оптимальному centruванні. Традиційними методами визначення центрів РБФ є [30]:

- випадковий вибір векторів з множини навчальних даних;
- застосування схем навчання із учителем.

У даній роботі пропонується підхід до визначення кількості та місця розташування центрів РБФ-мережі за допомогою алгоритму клонального відбору (АКВ). Крім того, додержуючись ідеї комплексного підходу до розв'язання задачі настроювання параметрів нейронної мережі, ми використовували АКВ в якості єдиного (глобального) інструмента пошуку оптимальних значень всіх параметрів, що настроюють. Нижче наводиться опис елементів АКВ, які повинні бути адаптовані для розв'язання поставленої задачі [32, 33, 34].

3.2. Опис конфігурації штучних імунних систем для розв'язання задачі настроювання параметрів РБФ-мережі

Однією з теорій утворення антитіл є клонально-селекційна теорія, розроблена Бернетом [37]. Основне положення її полягає в тому, що здатність індивідуума розпізнавати антиген пов'язана з певними імунологічно реактивними лімфоцитами або генетично ідентичними лініями лімфоцитів (клонами). Коли В-клітини розпізнають антиген, вони стимулюються, після чого клонуються і синтезують антитіла з тією же специфічністю. Цей процес стимулює тільки ті клітини, які синтезують корисний тип антитіл, і називається клональним відбором. Число клонів, вироблених лімфоцитом, пропорційно рівню його стимуляції. Клони не досконалі, але вони піддані соматичній мутації (яка характеризується високим рівнем мутації); у результаті їхні нащадки мають антитіла, що відрізняються від антитіл їхніх батьків. Ці нові В-Клітини можуть також зв'язуватися з антигенами і, якщо вони мають високу афінність до антигенів, вони у свою чергу будуть знову активізовані та клоновані. Рівень клонування клітини пропорційний її «придатності» до задачі: найбільш придатні клітини більшою мірою піддані реплікації. У цілому наявність підвищеної мутабельності генів імуноглобулінів у процесі онтогенезу і забезпечує різноманіття антитіл. По існуючим сьогодні представленнях, створення різноманіття антитіл забезпечується двома альтернативними механізмами: 1) кодуванням кожної специфічної послідовності молекул антитіл в окремих генах; 2) зберігання інформації у відносно невеликій кількості генів, а розмаїтість антитіл може бути наслідком мутацій і рекомбінацій. По теорії рекомбінацій розмаїтість генів, що кодують варіабельні ділянки ланцюгів імуноглобулінів, виникає в соматичних клітинах у результаті рекомбінації обмеженого числа генів. Соматична мутація гарантує достатню розмаїтість множини клонів, у той час як відбір забезпечує конкуренцію за антигени. У цілому процес (фактично Дарвіністський) відбору та диференціювання рецепторів В-Клітини, ведучий до розвитку

популяцій тих В-Клітин, які краще пристосовані розпізнавати певні антигенні детермінанти, називається *дозріванням афінності*.

Формально алгоритм клонального відбору можна представити як [32,35,36]:

$$CLONALG = (P^l, G^k, l, k, m_{Ab}, \delta, f, I, \tau, AG, AB, S, C, M, n, d), \quad (3)$$

де P^l – простір пошуку (простір форм); G^k – подання простору; l – довжина вектора атрибутів (розмірність простору пошуку); k – довжина рецептора антитіла; m_{Ab} – розмір популяції антитіл; δ – функція експресії; f – функція афінності; I – функція ініціалізації початкової популяції антитіл; τ – умова завершення роботи алгоритму; AG – підмножина антигенів; AB – популяція антитіл; S – оператор селекції; C – оператор клонування; M – оператор мутації; n – кількість кращих антитіл, що відбирають для клонування; d – кількість гірших антитіл, що підлягають заміні новими.

Алгоритм клонального відбору для розв'язання задач оптимізації та класифікації детально описані нами в роботах [35,36], тому в даній роботі обмежимося лише основними визначеннями, необхідними для розуміння роботи клонального алгоритму. Введемо необхідні визначення:

Визначення 1: Алгоритм клонального відбору ґрунтується на імітації принципу клонального відбору, що складається з механізмів: клонального відбору, клонального розширення та дозрівання афінності на основі соматичної гіпермутації.

Ґрунтуючись на **Визначенні 1** можна виділити два важливі принципи алгоритму клонального відбору.

Принцип 1: Рівень проліферації кожної імунної клітини пропорційний її афінності щодо антигену (*вище відносна афінність, більше потомства*)

Принцип 2: Мутація, перенесена кожною імунною клітиною під час репродукції, зворотно пропорційно афінності рецептора клітини до антигену (*вище відносна афінність, нижче мутація*)

Визначення 2: Під *класом* у даній роботі будемо розуміти категорію даного вектора властивостей. Ідентифікатор класу, як правило, використовується при обчисленні афінності антитіл, виступаючи в ролі фенотипу (вихідного значення) клітини.

Визначення 3: Під *рівень клонування розуміється* скалярне значення, використовуване для визначення числа клонів, які генеруються одним антитілом.

Визначення 4: *Вектором властивостей* є один екземпляр даних, представлений як послідовність (вектор) значень. Кожна позиція в даній послідовності представляє одне із властивостей, пов'язаних з даними. При цьому у кожній властивості є свій інтервал зміни значень.

Визначення 5: *Рівень гіпермутації* являє собою цілочислове значення, яке використовується для визначення числа клонів, що піддаються мутації, отриманих із клітин пам'яті.

Визначення 6: *Загальний рівень мутації* – параметр між 0 і 1, що вказує на ймовірність того, що дану властивість антитіла (вектора властивостей) буде видозмінено.

Визначення 7: *Жорсткість відбору (селективний тиск)* – цілочислове значення, що представляє собою розмір випадкового змагального пулу антитіл, з якого на кожному кроці відбору вибирається одне краще антитіло.

Теорія і алгоритм клонального відбору для розв'язання задач апроксимації та синтезу РБФ-нейронних мереж описані нами в роботі [36]. У даній роботі нами вперше запропонована методологія синтезу РБФ-нейронних мереж для синтезу нейронних мереж у вигляді бінарних класифікаторів. Синтез кожної окремої нейронної мережі для кожного класу здійснюється у відповідності з алгоритмом, який представлено на рис.2.

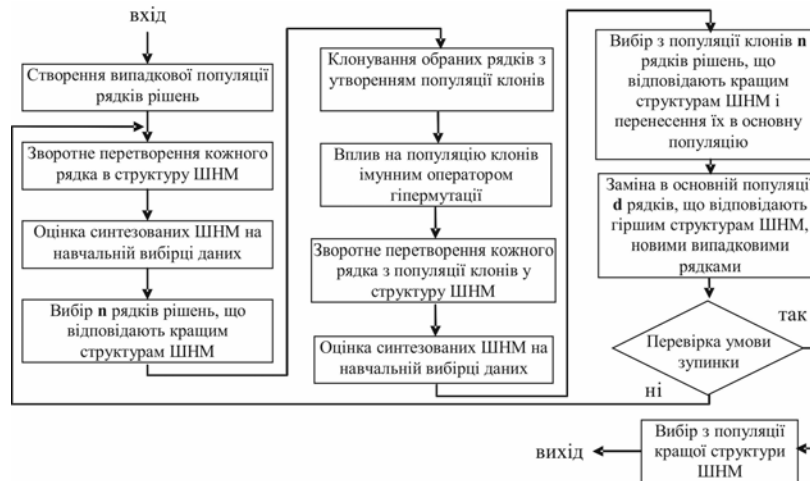


Рис. 2 Алгоритм синтезу РБФ-нейронної мережі алгоритмом клонального відбору

Грунтуючись на архітектурі нейронної мережі (рис. 1), в якості параметрів, що настраюються, можна виділити наступні:

f_1	c_{11}	...	c_{1n}	w_1	σ_1	...	f_m	c_{m1}	...	c_{mm}	w_m	σ_m	a
-------	----------	-----	----------	-------	------------	-----	-------	----------	-----	----------	-------	------------	-----

Рис. 3. Кодування параметрів, що настраюються, РБФ-нейронної мережі у вигляді антигіла (хромосоми):

- n : – кількість входів m ; – кількість нейронів f ; – вкл/викл нейрон (0 або 1)
- c – центри RBF; – w синаптичні ваги вихідного шару; σ – параметр RBF;
- a – параметр функції активації вихідного нейрона

У даній роботі досліджується синтез РБФ-нейронних мереж, що класифікують, при цьому класифікатором у загальному випадку називається функція, яка по вектору ознак об'єкта виносить рішення, якому саме класу він належить:

$$F : \mathfrak{R}^n \rightarrow Y \tag{4}$$

Функція F відображає простір векторів ознак у простір міток класів Y . У випадку двох класів $Y = \{0,1\}$, '1' відповідає випадку виявлення події, що шукається, '0' – подія не знайдена. Ми розглядаємо варіант навчання із учителем (supervised learning), коли для навчання класифікатора нам доступний деякий набір векторів $\{x\}$, для яких відома їхня істинна приналежність до одного із класів.

При бінарній класифікації ідентифікатори класів можна інтерпретувати як стани системи (активний або пасивний, нормальний або аномальний), які представлені рядом властивостей.

Визначення 8. Простір станів системи. Стан системи визначається вектором властивостей. Кожний стан $x^i = (x_1^i, \dots, x_n^i) \in [0,1]^n$ представлений множиною $U \subseteq [0,1]^n$. Елементи вектора властивостей можуть бути масштабовані або нормалізовані в інтервалі $[0,1]$.

Визначення 9. Нормальний підпростір. Множина векторів властивостей $Positiv \subseteq U$, що представляють нормальний стан системи. Його доповнення називають $Negativ$ і визначають як $Negativ = U - Positiv$. У деяких випадках ми будемо визначати множину $Positiv$ (або $Negativ$) використовуючи його характеристичну функцію $\chi_{Positiv} : [0,1]^n \rightarrow \{0,1\}$

$$\chi_{Positiv}(\vec{x}) = \begin{cases} 1 & \text{якщо } \vec{x} \in Positiv \\ 0 & \text{якщо } \vec{x} \in Negativ \end{cases} \tag{5}$$

Визначення 10. Задача бінарної класифікації РБФ-нейронною мережею. Для заданої множини позитивних прикладів $Positiv' \subseteq Positiv$ провести оцінку характеристичної функції

($\mathcal{X}_{Positiv}$) нормального простору, яка, у результаті, повинна мати здатність вирішувати, чи є спостережуваний стан системи нормальним або аномальним.

Вся множину нейронних мереж Δ можна розбити на зчисленну множину підкласів A , обумовлених обраною топологією РБФ-нейронної мережі (кількість РБФ-нейронів мережі Γ). У межах кожного класу $RBFN_i \subset RBFN$ нейронні мережі будуть характеризуватися додатковим набором параметрів: кількістю входів n , набором синаптичних ваг вихідного шару $W = \{w^i, i = 1, \dots, p\}$, кількістю РБФ-нейронів мережі $\Gamma = \{\gamma^i, i = 1, \dots, p\}$, центрами РБФ $C = \{c^i, i = 1, \dots, p\}$, параметрами РБФ $\Sigma = \{\sigma^i, i = 1, \dots, p\}$, параметром функції активації вихідного нейрона i -й мережі $A = \{\alpha^i, i = 1, \dots, p\}$. У такий спосіб формується вектор параметрів, що настроюються, РБФ-нейронної мережі $\theta = \{\Gamma, W, C, \Sigma, A\}$. Природним критерієм вибору РБФ-нейронної мережі буде функція, задана середньоквадратичним відхиленням для довільних вхідних даних.

Отже, задачу синтезу РБФ-нейронної мережі можна звести до задачі оптимізації типу

$$F^* = F(\theta^*) = \min F(\theta), \quad (6)$$

$$a_1 \leq x_1 \leq b_1, \dots, a_n \leq x_n \leq b_n,$$

де на функцію F не накладається ніяких обмежень, таких як диференційовність, умови Ліпшица, безперервність.

Для розв'язання задачі багатопараметричної оптимізації функції вигляду (6) використаємо відповідні оператори алгоритму клонального відбору.

Для кодування значень використана двійкова система. Точність представлення (кількість біт на одне значення) виділена як параметр настроювання ШС. Елементи рядка f_1, \dots, f_m кодують стани відповідних нейронів прихованого шару. Значення «0» відповідає пасивному або виключеному стану (нейрон не бере участь у розрахунку вихідного значення мережі). Значення «1» – нейрон активний (включений). Завдяки такій схемі забезпечується автоматичний пошук оптимальної кількості прихованих елементів РБФ-мережі; в настройках ШС встановлюється лише максимально можлива кількість цих елементів.

В якості цільової функції та функції афінності обрана середньоквадратична помилка мережі на навчальних даних. Навчальний алгоритм має наступні особливості реалізації:

- селекція реалізована за принципом турнірного відбору [39], що дає можливість керувати збіжністю алгоритму та підтримувати розмаїтість популяції на необхідному рівні;
- внаслідок специфіки двійкового кодування, запропонована схема мутації, відповідно до якої ймовірність зміни окремого біта рядка залежить не тільки від афінності антитіла в цілому, але і від значимості цього біта.

Розглянемо приклад, представлений на рис. 4.

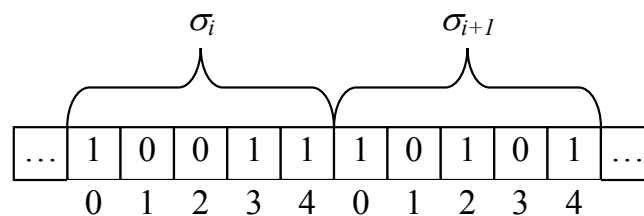


Рис. 4. Ділянка рядка антитіла із двійковим кодуванням

У даному прикладі мутація бітів з номерами 0 або 1 очевидно викличе більш істотні зміни обраних параметрів, чим мутація біта з номером 4. Коли індивідуум досягає досить високої афінності, більш значимі біти повинні поступово виключатися з операції мутації. Цього можна досягти, зменшуючи відповідні ймовірності застосування оператора до даних бітів. Формально цей процес можна зобразити так:

$$p_i = p_m \cdot \left(aff + \frac{i}{l} (aff_{\min} - aff) \right). \quad (7)$$

де p_i – імовірність мутації i -го біта параметра в рядку антитіла; p_m – загальний рівень мутації, що задається як параметр алгоритму; aff – поточна афінність антитіла; l – точність (кількість біт) подання кодуємих параметрів; aff_{\min} – мінімальне значення афінності, що відповідає обраній цільовій функції та функції афінності.

Типи базисних функцій і активаційної функції лінійного шару задаються як параметри ШС.

Навчання і синтез колективів нейронних мереж здійснюється за схемою, представленої на рис. 5.

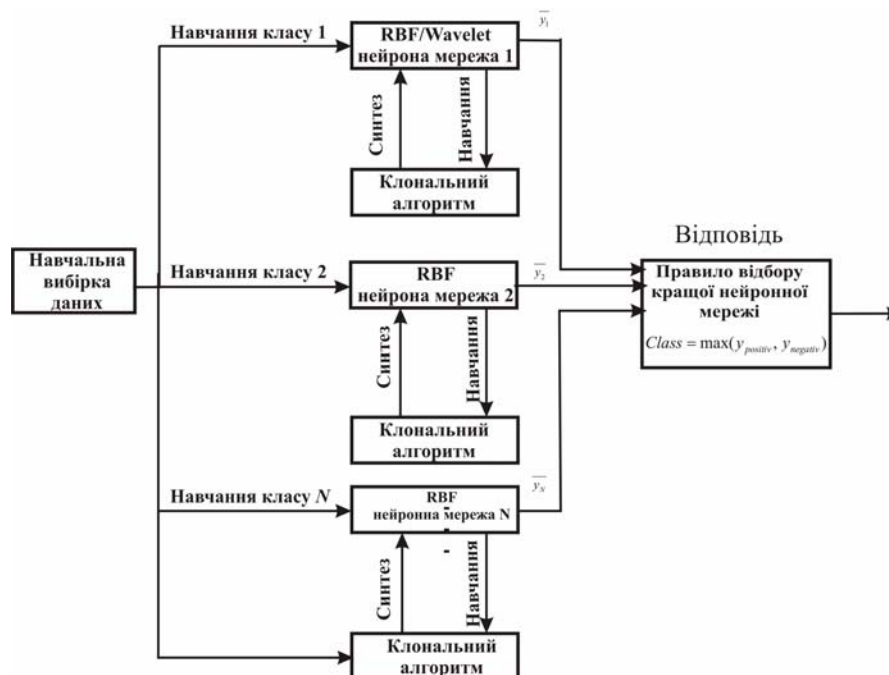


Рис. 4. Синтез колективу бінарних класифікаторів РБФ-нейронних мереж

Ідентифікація стану системи здійснюється за виразом (8):

$$Class = \max(y_{\text{positiv}}, y_{\text{negativ}}) \quad (8)$$

4. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

4.1. Опис даних

Задача 1. Класифікація у випадку двох класів, що перекриваються (Clouds database, ELENA Project). Ціль даної задачі – побудова класифікатора для двох класів, що сильно перекриваються, з нелінійною межею, що розділяє класи. Вихідні дані (2 входи – дійсні числа, 1 вихід – номер класу) містяться у файлі. Файл містить 5000 прикладів (приклади одного класу зображені дрібними крапками, другого – великими). Для навчання використовуються перші 4000 прикладів, інші – для тестування. Потрібно побудувати нейронну мережу, що дає мінімальну помилку при тестуванні.

Задача 2. Класифікація у випадку двох класів, що не перекриваються (Concentric database, ELENA Project). В цій задачі досліджується побудова нейромережного класифікатора для двох класів, що не перекриваються, які є лінійно нероздільними. Вихідні дані (2 входи – дійсні числа, 1 вихід – номер класу) містяться у файлі. Файл містить 2500 прикладів (приклади одного класу зображені дрібними крапками, другого – великими). Для навчання використовуються перші 1800 прикладів, інші – для тестування. Потрібно побудувати нейронну мережу, що дає мінімальну помилку при тестуванні.

4.2. Оцінка впливу мутації на збіжність

У роботі досліджувався вплив рівня мутації на збіжність алгоритму. Загальний рівень мутації послідовно відбирався: 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09. Для навчання мережі використовувалося 500 ітерацій. Як міра якості оцінки класифікатора використовувалося середньоквадратичне відхилення:

$$RBFN = E \left\{ \frac{1}{N} \sum_{i=1}^N [y_i^* - f(x_i^*)]^2 \right\} \quad (9)$$

Тут (\bar{x}_i^*, y_i^*) – точки тестової вибірки які не беруть участь у навчанні при конструюванні нейронної мережі. Функція $E\{\cdot\}$ визначає дисперсію аргументу. Оцінка $RBFN$ характеризує узагальнюючу здатність конкретної нейронної мережі.

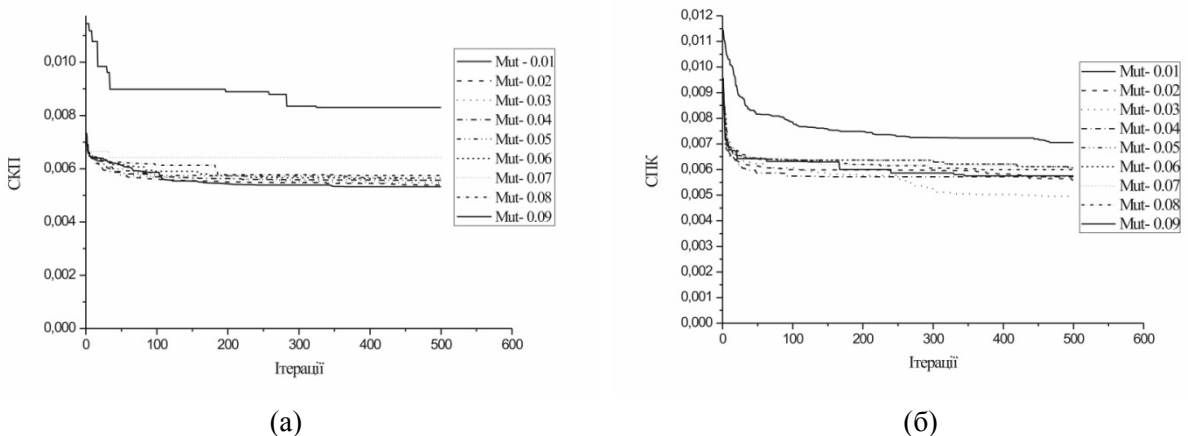


Рис. 5. Збіжність алгоритму при різних рівнях мутації:
 а) класифікація у випадку двох класів, що перекриваються (Clouds);
 б) класифікація у випадку двох класів, що не перекриваються (Concentric)

4.3. Оцінка класифікаторів

Для оцінки отриманих колективів бінарних нейромережних класифікаторів використовувалися ROC-Криві, AUC – чисельний показник площі під кривою і % правильно розпізнаних об’єктів на тестовій множині.

ROC-Крива (Receiver Operator Characteristic) являє собою криву, яка використовується для представлення результатів бінарної класифікації в машинному навчанні. Один клас називається класом з позитивними наслідками, другий – з негативними наслідками. ROC-Крива показує залежність кількості вірно класифікованих позитивних прикладів від кількості невірно класифікованих позитивних прикладів. В термінології ROC-Аналізу перші називаються істинно позитивною, другі – хибно позитивною множиною. При цьому передбачається, що в класифікатора є деякий параметр, варіюючи який, ми будемо одержувати ту або іншу розбивку на два класи. Цей параметр часто називають порогом, або точкою відсікання (cut-off value). Залежно від нього будуть утворюватися різні величини помилок I і II роду. Для розуміння суті помилок I і II роду розглянемо чотирипільну таблицю спряженості (confusion matrix), що будується на основі результатів класифікації моделлю і фактичної (об’єктивної) приналежності прикладів до класів.

Таблиця 1

Таблиця спряженості моделлю і фактичної приналежності прикладів до класів

Модель	Фактично	
	позитивно	негативно
позитивно	TP	FP
негативно	FN	TN

- TP (True Positives) – вірно класифіковані позитивні приклади (так звані істинно позитивні випадки);
- TN (True Negatives) – вірно класифіковані негативні приклади (істинно негативні випадки);
- FN (False Negatives) – позитивні приклади, класифіковані як негативні (помилка I роду). Це так званий «помилковий пропуск» – коли подія, що нас цікавить, помилково не виявляється (хибно негативні приклади);
- FP (False Positives) – негативні приклади, класифіковані як позитивні (помилка II роду); Це помилкове виявлення, тому що при відсутності події помилково виноситься рішення про її присутність (хибно позитивні випадки).

При аналізі частіше оперують не абсолютними показниками, а відносними – частками (rates):

$$\text{Частка істинно позитивних прикладів (True Positives Rate): } TPR = \frac{TP}{TP + FN}$$

$$\text{Частка хибно позитивних прикладів (False Positives Rate): } FPR = \frac{FP}{TN + FP}$$

$$\text{Чутливість (Sensitivity) – частка істинно позитивних випадків: } Se = TPR = \frac{TP}{TP + FN}$$

Специфічність (Specificity) – частка істинно негативних випадків, які були правильно ідентифіковані моделлю: $Sp = \frac{TN}{TN + FP}$

Модель із високою чутливістю часто дає істинний результат при наявності позитивного результату (виявляє позитивні приклади). Навпаки, модель із високою специфічністю частіше дає істинний результат при наявності негативного результату (виявляє негативні приклади). ROC-Крива утворюється у такий спосіб:

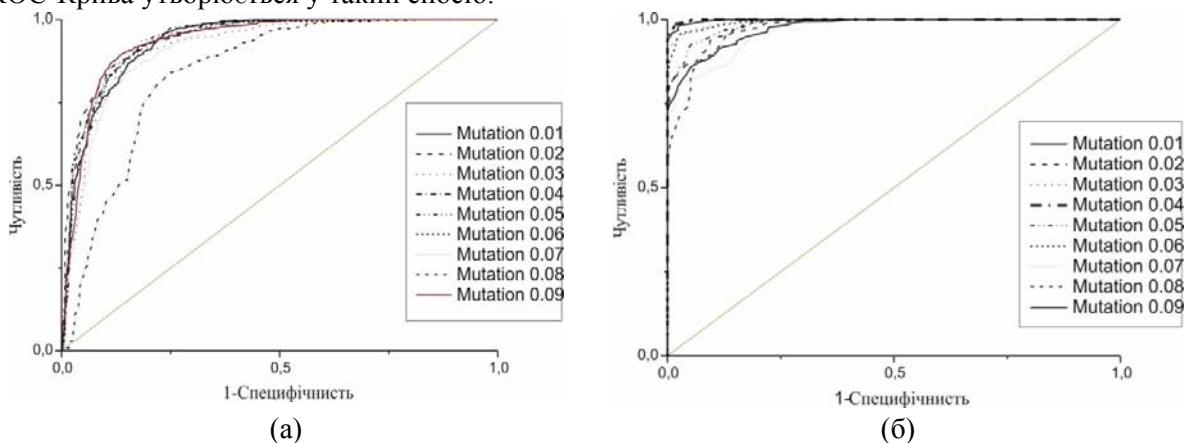


Рис. 6. Крива ROC при різних рівнях мутації:

- а) класифікація у випадку двох класів, що перекриваються (Clouds);
 б) класифікація у випадку двох класів, що не перекриваються (Concentric)

Для кожного значення порога відсікання, що змінюється від 0 до 1 із кроком dx (наприклад, 0.01), розраховуються значення чутливості Se і специфічності Sp . Як альтернатива порогом може бути кожне наступне значення прикладу у вибірці.

Будується графік залежності: по осі Y відкладається чутливість Se , по осі X – $100\% - Sp$ (сто відсотків мінус специфічність), або, що та ж саме, FPR – частка хибно позитивних випадків.

У результаті дослідження нами були отримані наступні криві (рис. 6).

Для ідеального класифікатора графік ROC-Кривої проходить через верхній лівий кут, де частка істинно позитивних випадків складає 1.0 (ідеальна чутливість), а частка хибно позитивних прикладів дорівнює нулю. Тому чим ближче крива до верхнього лівого кута, тим вище передбачувальна здатність моделі. Навпаки, чим менше вигин кривої і чим ближче вона

розташована до діагональної прямої, тим менш ефективна модель. Діагональна лінія відповідає «марному» класифікатору, тобто повної нерозрізненості двох класів. Візуальне порівняння кривих ROC не завжди дозволяє виявити найбільш ефективну модель. Своєрідним методом порівняння ROC-Кривих є оцінка площі під кривими. Теоретично вона змінюється від 0 до 1.0, але, оскільки модель завжди характеризується кривою, розташованою вище позитивної діагоналі, то звичайно говорять про зміни від 0.5 («марний» класифікатор) до 1.0 («ідеальна» модель). Ця оцінка може бути отримана безпосередньо обчисленням площі під багатогранником, обмеженим праворуч і знизу осями координат і ліворуч угорі – експериментально отриманими точками. Чисельний показник площі під кривою називається AUC (Area Under Curve). Обчислити його можна, наприклад, за допомогою чисельного методу трапецій:

$$AUC = \int f(x)dx = \sum_i \left[\frac{X_{i+1} + X_i}{2} \right] * (Y_{i+1} - Y_i) \quad (10)$$

Прийнято вважати, що чим більше показник AUC, тим кращою прогностичною силою володіє модель. Однак варто знати, що:

- показник AUC призначений скоріше для порівняльного аналізу декількох моделей;
- AUC не містить ніякої інформації про чутливість і специфічність моделі.

У літературі іноді наводиться наступна експертна шкала для значень AUC, по якій можна судити про якість моделі:

Інтервал AUC	Якість моделі
0.9-1.0	Відмінне
0.8-0.9	Дуже гарне
0.7-0.8	Гарне
0.6-0.7	Середнє
0.5-0.6	Незадовільне

Ідеальна модель має 100% чутливість і специфічність. Однак на практиці домогтися цього неможливо, більше того, неможливо одночасно підвищити і чутливість, і специфічність моделі.

Таблиця 2

Результати класифікації

№	Рівень мутації	% правильно розпізнаних об'єктів на тестовій вибірці <i>Clouds</i>	% правильно розпізнаних об'єктів на тестовій вибірці <i>Concentric</i>	AUC (Clouds)	AUC (Concentric)
1	0.01	84.62	96.71	0,9338	0,99722
2	0.02	84.22	94.71	0,9445	0,99694
3	0.03	84.62	96.57	0,91964	0,99864
4	0.04	83.82	96.43	0,93657	0,99656
5	0.05	84.62	95.29	0,93834	0,99751
6	0.06	81.62	96.14	0,93599	0,99757
7	0.07	81.02	90.71	0,91557	0,99565
8	0.08	71.63	94.29	0,84313	0,99555
9	0.09	83.32	92.57	0,93549	0,99371

ВИСНОВКИ

1. Розроблено новий комбінований метод бінарної класифікації, який базується на синтезі колективу РБФ-нейронних мереж за допомогою штучних імунних систем.

2. Проведено дослідження та аналіз впливу загального рівня мутації на якість бінарної класифікації синтезованих нейронних мереж радіального базису.
3. Наведено результати обчислювальних експериментів, що показують ефективність запропонованого комбінованого методу.

ЛІТЕРАТУРА

1. Hrycej T. Modular learning in neural networks: a modularized approach to neural network classification. – Wiley, 1992. – P. 256.
2. Ferdinando D., Calabretta A., Parisi D. Evolving modular architectures for neural networks // Proc. of the Sixth Neural Computation and Psychology Workshop Evolution, Learning and Development. – London: Springer Verlag. – 2001. – P. 253-262.
3. Rojas R. Neural Networks. A Systematic Introduction. – Berlin: Springer-Verlag, 1996. – 502 p.
4. Pham D. T., Liu X. Neural Networks for Identification, Prediction and Control. – London: Springer-Verlag, 1995. – 238 p.
5. Haykin S. Neural Networks. A Comprehensive Foundation. – Upper Saddle River, N.J.: Prentice Hall, Inc., 1999. – 842 p.
6. Hansen L. K., Salamon P. Neural networks ensembles // IEEE Trans. on Pattern Analysis and Machine Intelligence. – 1990. – 12. – P. 993 – 1000.
7. Sharkey A. J. C. On combining artificial neural nets // Connect. Sci. – 1996. – 8. – P. 299 – 313.
8. Hashem S. Effects on collinearity on combining neural networks // Connect. Sci. – 1996. – 8. – P. 315 – 336.
9. Opitz D. W., Shavlik J. W. Actively searching for an effective neural network ensemble // Connect. Sci. – 1996. – 8. – P. 337 – 353.
10. Hashem S. Optimal linear combination of neural networks // Neural Networks. – 1997. – 10. – P. 599-614.
11. Naftaly U., Intrator N., Horn D. Optimal ensemble averaging of neural networks // Network: Comput. Neural Syst. – 1997. – 8. – P. 283-296.
12. Giacinto G., Roli F. Design of effective neural network ensembles for image classification purposes // Image and Vision Computing, Elsevier Science. – 2001. – Vol. 19. – P. 699-707.
13. Kussul M., Galinskaya A. Comparative study of modular classifiers and its training // Proc. of the X-th International Conference “Knowledge – Dialog – Solution” KDS 2003, Varna, Bulgaria. – 2003. – 16-26 June. – P.168-174.
14. Sharkey A.J.C. Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems. – London: Springer-Verlag, 1999. – P. 298.
15. Jacobs R. A., Jordan M. I., Nowlan S., Hinton G. E. Adaptive mixtures of local experts // Neural Computation. – 1991. – Vol. 3. – P. 1-12.
16. Wanas N., Kamel M. Decision fusion in neural network ensembles // International Joint Conference on neural Networks (IJCNN'01). – Washington D.C., USA. – 2001. – Jul 15-19. – Vol. 4. – P. 2952-2957.
17. Mak M.W. Speaker Identification using Modular Recurrent Neural Networks // Proc. 4th IEE International Conf. on Artificial Neural Networks – UK. – 1995. – June. – P. 1-6.
18. Silva C., Crisóstomo M., Ribeiro B. A Modular Neural Architecture for Navigating NOMAD Mobile Robot // 8th International Conf. on Information Processing and Management of Uncertainty in Knowledge Based Systems. – Spain. – 2000. – 3-7 July. – P. 1199-1204.
19. Giacinto G., Roli F. Ensembles of Neural Networks for Soft Classification of Remote Sensing Images // Proc. of the European Symposium on Intelligent Techniques. – Bari, Italy. – 1997. – March 20-21. – P. 166-170.
20. Zhou Z. -H., Jiang Y., Yang Y. -B., Chen S. -F. Lung cancer cell identification based on artificial neural network ensembles // Artificial Intelligence in Medicine. – 2002. – Vol. 24(1). – P. 25-36.
21. Бодянский Е.В., Руденко О.Г. Искусственные нейронные сети: архитектуры, обучение, применение // Харьков: ТЕЛТЕХ 2004, – 372 с.
22. Hüsken M., Igel C., Toussaint M. Task-Dependent Evolution of Modularity in Neural Networks – A Quantitative Case Study // Genetic and Evolutionary Computation Conference (GECCO 2001). – Late Breaking Papers. – 2001. – P. 187-193.
23. Roli F., Kittler J., Fumera G., Muntoni D. An Experimental Comparison of Classifier Fusion Rules for Multimodal Personal Identity Verification Systems.- Multiple Classifier Systems. – 2002. – P. 325-336
24. Бардачев Ю.Н., Гринавцев О.В., Литвиненко В.И., Фефелов А.А. Синтез нечетких нейронных сетей с помощью иммунных алгоритмов для прогнозирования электрической нагрузки // Збірник наукових праць «Моделювання та керування станом еколого-економічних систем регіону». Випуск 3, 2006, с. 123-133.
25. Гринавцев О.В., Фефелов А.А., Бидюк П.И., Литвиненко В.И. Метод коллективного принятия решений при решении задач прогнозирования с помощью нейронных сетей // Тезисы докладов Ежегодной научно-технической конференции молодых ученых и специалистов ИПМ им. Г.Е. Пухова. – 2006 г.
26. Фефелов А.А., Горавский С.П., Литвиненко В.И. Построение и настройка вейвлет-нейронных сетей при помощи искусственных иммунных систем // Збірник наукових праць у п'яти томах Міжнародної наукової конференції Інтелектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій/ Євпаторія 2005, Том 5, с.77-80.
27. Медведев В.С., Потемкин В.Г. Нейронные сети. MATLAB 6 / Под общ. ред. к.т.н. В.Г. Потемкина. – М.: ДИАЛОГ-МИФИ, 2002. – 496с.

28. Назаров А.В., Лоскутов А.И. Нейросетевые алгоритмы прогнозирования и оптимизации систем – СПб.: Наука и Техника, 2003. – 384 с.: ил.
29. Нейронные сети. STATISTICA Neural Networks // Пер. с англ. – М.: Горячая линия – Телеком, 2001. – 182 с.
30. Руденко О.Г., Бодянский Е.В. Основы теории искусственных нейронных сетей. – Харьков: ТЕЛТЕХ, 2002.–317 с.
31. De Castro L.N., Von Zuben F.J. An Immunological Approach to Initialize Centers of Radial Basis Function Neural Networks. (Pre-print). Proc. Of CBRN'01 (Brazilian Conference on Neural Networks), 2001, pp. 79-84.
32. Бидюк П.И., Литвиненко В.И., Молчанов В.Ю., Фефелов А.А. Методы определения повреждения и отказов систем управления с помощью нейронных сетей // Збірник наукових праць Севастопольського військово-морського ордена Червоної Зірки інституту ім. П.С. Нахімова // Випуск 2(8) 2005 с.270-277.
33. Фефелов А.А., Бидюк П.И., Горавский С.П., Литвиненко В.И. Применение нейронных сетей для диагностики отказов технических систем // Тези доповідей Третьої міжнародної науково-практичної конференції Математичне та програмне забезпечення інтелектуальних систем (MPZIS-2005) 16-18 листопада 2005 року, м. Дніпропетровськ, с.11-12.
34. Фефелов А.А., Бидюк П.И., Литвиненко В.И. Применение комбинированных искусственных иммунных систем при решении задач построения моделей динамических процессов // Міжнародний семінар з індуктивного моделювання. Збірник праць // Відп. Редактор д.т.н. Степашко В.С.– К.: Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, 2005. – С. 69-84.
35. Литвиненко В.И., Фефелов А.А., Горавский С.П. Объектно-ориентированная реализация алгоритма клональной селекции // Радіоелектроніка. Інформатика. Управління / Запоріжжя, 2003 (9). – С. 81-88
36. Литвиненко В.И. Экспериментальные исследования алгоритма клонального отбора при решении задач оптимизации многоэкстремальных функций // Вестник ХНТУ. – 2007. – № 3 (29). – С. 29-41.
37. Синтез радиально-базисных сетей для решения задачи дистанционного определения концентрации хлорофила // ААЭКС – 2007 №11(19). с.53-65
38. Burnet, F.M.. A modification of Jerne's theory of antibody production using the concept of clonal selection. *Australian Journal Science* 1957 №20 p67–69. 37.Литвиненко В.И.
39. Скурихин А.Н. Генетические алгоритмы //Новости искусственного интеллекта, 1995, № 4, с.6-46.

Рецензенти: д.т.н., проф. Кондратенко Ю.П.
 д.т.н., проф. Фісун М.Т.

© Литвиненко В.І., Фефелов А.О.,
 Дідик О.О., 2009

Стаття надійшла до редколегії 21.03.09