

УДК 004.94

Моделювання процесу оброблення двовимірного масиву
даних за різницевиими зрізами

Мартинюк Т.Б., Хом'юк В.В., Куперштейн Л.М.

Вінницький національний технічний університет

Анотація. Розглянуто особливості оброблення за методом різницевиих зрізів двовимірного масиву даних розмірністю $m \times n$, який сформовано з n зважених елементів вхідного вектора для m дискримінантних функцій. Отримано результати імітаційного моделювання наведеного процесу оброблення двовимірного масиву даних.

Summary. The features of treatment after the method by difference slices of two-dimensional data array are considered a dimension $m \times n$, which is formed from the n weighed elements of entrance vector for m discriminant function. The results of imitation design of the resulted process of treatment of two-dimensional data array are got.

Одним з відомих методів розпізнавання образів є їх класифікація з використанням дискримінантних функцій (ДФ) як розподільчих функцій у просторі ознак для кожного класу [1]. Особливо ефективно цей підхід використовується у медичній діагностиці, оскільки існує можливість формування коефіцієнтів лінійних ДФ методами дискримінантного аналізу за певною обмеженістю навчальної вибірки [2]. В цьому випадку коефіцієнти ДФ представляють собою за аналогією з нейромережним підходом [3] вагові коефіцієнти при навчанні з «вчителем» [1]. Результатом класифікації за ДФ є визначення максимальної серед усіх ДФ, місцезнаходження (позиція) якої визначає клас, до якого належить вхідний образ, поданий вектором його ознак [2,3].

В даній роботі досліджується новий підхід до визначення максимальної ДФ, який дозволяє не тільки відмовитись від формування ДФ як суми зважених елементів вхідного вектора ознак, але й значно розширити функціональні можливості таких класифікаторів. Наприклад, разом з формуванням вихідного вектора належності вхідного образу до певного класу існує можливість сформувати вектор рангів ДФ, що дозволить визначити всі ДФ, найближчі до максимальної [4].

В цьому випадку зважені елементи (доданки) всіх ДФ розглядаються як двовимірний масив даних, який обробляється за методом різницевиих зрізів (РЗ) [5]. Особливістю такого підходу є

імовірнісна залежність часових показників процесу оброблення не тільки від розмірності двовимірного масиву даних, але й від закону розподілу його елементів. Отже, метою даної роботи є дослідження залежності часових показників від вигляду двовимірного масиву даних для формування математичної моделі цієї залежності.

Алгоритм оброблення двовимірного масиву даних за методом РЗ з формуванням вектора належності вхідного образу до певного класу має такий вигляд [6].

Початкові дані: вхідний образ у вигляді n – вимірний вектора ознак $Z = \{z_1, \dots, z_n\}$; матриця ваг (коефіцієнтів ДФ) $W = |w_{ij}|, i = \overline{1, m}, j = \overline{1, n}$, де m – кількість класів; n – розмірність вектора ознак; m – вимірний вихідний вектор $Y = \{y_1, \dots, y_m\}$, причому спочатку $y_i = 0$; матриця A^0 елементів всіх ДФ $g_i(Z)$ вигляду

$$g_i(Z) = \sum_{j=1}^n w_{ij} \cdot z_j, \quad (1)$$

причому

$$a_{ij}^0 = w_{ij} \cdot z_j. \quad (2)$$

Тоді i – й рядок A_i^0 матриці A^0 , який складається з елементів a_{ij}^0 (2) i – ої ДФ $g_i(Z)$ в подальшому розглядається як i – й масив чисел.

Крок 1. Введення двовимірного масиву A^0 .

Крок 2. Пошук мінімального елемента в кожному стовпці масиву $A^t, t = \overline{0, N}$, де N – кількість циклів оброблення, не враховуючи кількість рядків, всі елементи яких дорівнюють нулю, та формування з них вектор-рядка.

Крок 3. Паралельне віднімання від кожного стовпця масиву A^t відповідного елемента, отриманого на попередньому кроці вектор-рядка.

Крок 4. Перевірка умови появи поточного нульового рядка $A_k^t, k = \overline{1, m}$ та умови наявності m нульових рядків. При виконання другої умови перехід до кроку 6, інакше перехід до кроку 5.

Крок 5. Переміщення не рівних нулю елементів кожного рядка у крайню ліву вільну позицію. Перехід до кроку 2.

Крок 6. Формування вихідного вектора $Y = \{y_1, \dots, y_m\}$, в якому тільки один елемент y_l дорівнює одиниці, що відповідає останньому обнуленому l -му рядку матриці $\overline{\mathbf{A}}^N$ і свідчить про належність вхідного образу до l -го класу, $l = \overline{1, m}$.

Послідовність кроків 2-5 складає цикл оброблення.

В процесі моделювання використовувались двовимірні масиви даних, елементи яких згенеровані за нормальним законом розподілу. В експерименті значення σ змінювалося з кроком 10 від 120 до 10, при цьому розглядалось по 100 матриць відповідної розмірності. Програмна реалізація проводилась на мові C++ в програмному середовищі C++Builder 5. Отримані результати (середня кількість циклів оброблення) представлено рис.1, причому для випадку двовимірного масиву за умови $m < n$ отримані результати представлені на тривимірній стовпчастій діаграмі.

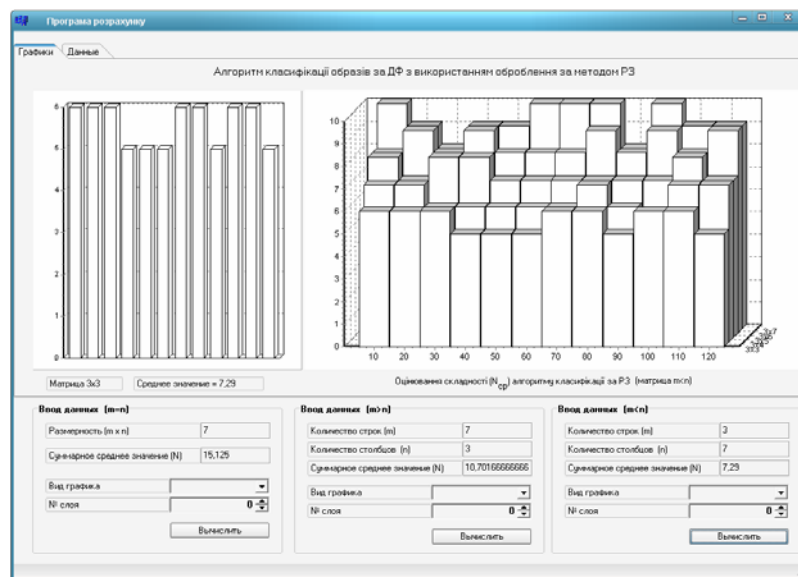


Рис.1 Комп'ютерне моделювання процесу оброблення двовимірного масиву даних за РЗ

Аналіз залежності кількості циклів $N_{\text{п0}}$ від розмірності двовимірного масиву даних свідчить про нелінійне зростання часу

оброблення при збільшені розмірності самих масивів, при цьому оброблення квадратних матриць ($m = n$) займає більше часу, ніж матриць, де $m \neq n$. Як видно з рис. 1 в процесі оброблення присутні «провали» кількості циклів оброблення як при різних значеннях σ , так і при різних розмірностях двовимірних масивів даних. Це, на думку авторів, відбувається в результаті зростання кількості однакових елементів у числовому масиві даних, тобто відбувається «скупчення» елементів навколо певних елементів даного масиву.

Аналіз результатів моделювання процесу оброблення двовимірного масиву даних за РЗ дає підставу стверджувати, що за умови, коли у вхідному масиві даних присутні групи однакових елементів, відбувається збільшення швидкодії процесу оброблення, що, звичайно, потребує додаткових досліджень. Але вже можна сказати, що перспективними областями застосування алгоритму оброблення двовимірного масиву даних за методом РЗ є такі, де обробляються двовимірні масиви даних із групами однакових елементів.

Список літератури

1. Хайкин С. Нейронные сети: полный курс/С. Хайкин; пер. с англ. – 2-е изд. – М.: ООО «И.Д Вильямс», 2006. – 1104 с. – ISBN 5 – 8459 – 0890 – 6.
2. Юнкеров В.И. Математико-статистическая обработка данных медицинских исследований/В.И. Юнкеров, С.Г. Григорьев. – СПб.: ВМедА, 2002. – 266 с. – ISBN 5 – 94277 – 011 – 5.
3. Бернюков А.К. Распознавание биоэлектрических сигналов/А.К. Бертоков, Л.Т. Сушкова//Зарубежная радиоэлектроника. – 1996. – № 12 – С. 47 – 51.
4. Патент України 24622, МПК 8 G06K9/00. Пристрій для класифікації образів/ Т.Б. Мартинюк, Ж.О. Бітюкова, Козлова В.І., М.В. Тончанюк; Вінниц. нац. техн. ун-т. – № 200701228; заявл. 05.02.2007; опубл. 10.07.2007, Бюл. № 10.
5. Мартинюк Т.Б. Рекурсивні алгоритми багатооперандної обробки інформації/Т.Б. Мартинюк. – Вінниця: УНІВЕРСУМ – Вінниця, 2000. – 216 с. – ISBN 966 – 7199 – 98 – 3.
6. Мартынюк Т.Б. Классификатор биоэлектрических сигналов/Т.Б. Мартынюк, А.Г. Буда, В.В. Хомюк, А.В. Кожемяко, Л.М. Куперштейн//Искусственный интеллект. – 2010. – № 3 – С. 88 – 95. – ISBN 1561 – 5359.