

МОДЕЛЮВАННЯ ПРОЦЕСУ РАНЖИРУВАННЯ ЗНАЧЕНЬ ДИСКРИМІНАНТНИХ ФУНКЦІЙ

Медвідь А.В.

Науковий керівник – доц., к.т.н. Мартинюк Т.Б.

Найпоширеніше застосування дискримінантний аналіз (ДА) знаходить у медичній діагностиці при класифікації (діагностуванні) захворювань за ознаками (симптомами) хвороби. Це пов'язано з тим, що використання лінійних методів ДА, наприклад, за дискримінантними функціями (ДФ), дозволяє зробити попередній діагноз на обмежених статистичних даних біомедичних досліджень. Разом з тим, при попередньому діагностуванні бажано визначити не тільки один найбільш вірогідний діагноз, але й найближчі до нього за симптомами.

Ранжирування у загальному випадку є результатом впорядкування елементів масиву і найчастіше виконується в процесі сортування масивів даних, що є складною процедурою. Застосування методу різницевого зрізів (РЗ) для визначення серед m ДФ максимальної $\max g_l(X)$, де $l = \overline{1, m}$ дозволяє, крім класифікації, також провести і процес ранжирування.

Суть ранжирування m ДФ $g_i(X)$ при такому підході полягає у наступному: для визначення рангів відповідних ДФ $g_i(X)$ необхідно ввести вихідний вектор рангів $R = \{r_1, \dots, r_m\}$, всім елементам r_i якого на початку оброблення присвоюються одиничні значення на відміну від вектора класифікації $P = \{p_1, \dots, p_m\}$, всім елементам p_i якого на початку оброблення присвоюються нульові значення. При цьому послідовне збільшення на одиницю (інкремент) певних елементів вектора рангів R після кожного обнуління відповідного рядка A_i^t поточної матриці A^t приведе до того, що по завершенню процесу класифікації сформується вектор рангів R , елементи якого відповідають відсортованим рядкам A_i^0 початкової матриці A^0 , що відповідає ранжируванню ДФ $g_i(X)$. При цьому мінімальній ДФ відповідає ранг, що дорівнює одиниці, а максимальній ДФ – ранг, що дорівнює m .

Наведену модель реалізовано на мові програмування MATLAB. Проведено три типи експериментів в залежності від способу подання елементів початкової матриці A^0 : а) з дробовими значеннями приведених елементів; б) з округленням до цілих значень елементів; в) з домноженням на 10 елементів. Останній експеримент обумовлений необхідністю реалізації досліджуваного процесу оброблення двовимірних масивів даних на програмованих логічних ІС (ПЛІС) з поданням вхідних даних цілими числами зі збереженням точності (один знак після коми).