

*О.Ю. Гончарук, О.В. Монченко, канд. техн. наук, доц.,
(Національний університет охорони здоров'я
України ім. П.Л. Шупика, Україна)
К.О. Безвершинюк, С.Г. Корчева
(Національний авіаційний університет, Україна)*

Використання критерію мінімуму відстані в задачах оптимізації комплексної терапії пацієнтів

У роботі було проаналізовано загальні проблеми комплексної терапії. Так, в дослідженні приймали участь три контрольні групи пацієнтів з артеріальною гіпертензією та проблемами ожирінням. Проаналізувавши отримані дані, було запропоновано використання критерію мінімуму відстані Махаланобіса між групами пацієнтів, що дозволило проаналізувати ефективності різних методик лікування та вибрати оптимальний метод комплексної терапії.

Дослідження присвячені обрахунку методик лікування надлишкової ваги. Дослідження проводили в період з 2017 по 2021 роки на клінічних базах кафедри сімейної медицини та амбулаторно-поліклінічної допомоги НМАПО імені П. Л. Шупика.

Загальна кількість обстежених осіб склала 149, з них 128 пацієнти з артеріальною гіпертензією (АГ) та ожирінням і 21 обстежений - група відносно здорових осіб. Пацієнти за АГ та ожирінням були розподілені на 2 групи, рандомізовані за віком, статтю та коморбідною патологією.

Проведенні дослідження були засновані на класифікації образів, спираючись на один з найпростіших і евристичних підходів – використання для класифікації образів функцій відстані. Вибір функцій відстані в якості інструмента класифікації є результатом того, що найбільш очевидний спосіб введення міри схожості для векторів образів, які можна розглядати як точку в евклідовому просторі - визначення їх близькості.

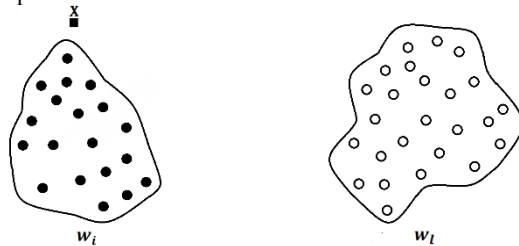


Рис.1. Образи, які піддаються класифікації, за допомогою поняття близькості

З рис .1. можна побачити, що вектор X належить до класу ω_i виключно з тих міркувань, що цей вектор знаходиться ближче до векторів образу ω_i . Можна

розраховувати на отримання позитивних практичних результатів при класифікації образів за допомогою функції відстані тільки в тих випадках, коли класи образів проявляють тенденцію до появи кластеризаційних властивостей. Цю особливість можна оцінити співставивши рис. 1 і рис. 2. Вивчення рис.1 показує, що віднесення вектора X до класу ω_i не викликає сумнівів. Якщо подивитись на рис.2, то доволі важко сказати, до якого класу належить вектор X , хоч класи і не перетинаються.

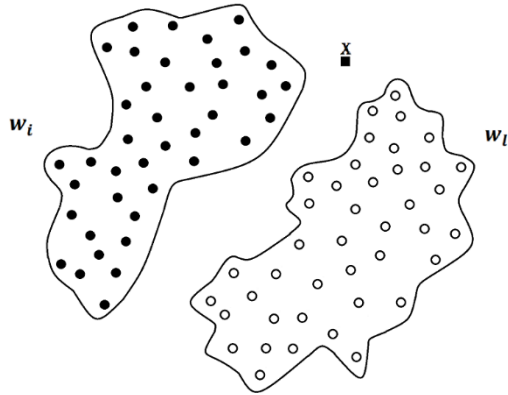


Рис.2. Образи, класифікація яких за допомогою поняття близькості викликає складнощі

Загальна оцінка таких ідей будуватиметься на рівні відповідної математичної строгості. Близькість образу, який ми будемо класифікувати, до образу певного класу – еталона, буде використовуватись в якості критерію його класифікації. Такий підхід називається класифікацією образів за критерієм мінімуму відстані [1].

Важливим моментом є вибір еталона або центра класера, який буде відігравати головну роль в побудові класифікатора образів за принципом мінімуму відстані. В представленому в роботі випадку таким центром класера є вибірка здорових людей. Мірою схожості (подібності), яка буде покладена в основу вимірювання відстані між образами, була застосована відстань Махаланобіса[1].

В загальному випадку задачу класифікації можна звести до задачі перевірки гіпотез [1,2], яка полягає в прийнятті однієї з K взаємно несумісних гіпотез H_0, H_1, \dots, H_k щодо невідомих характеристик об'єкту досліджень по результатах оцінювання багатомірного вектора $E(\lambda_1, \lambda_2, \dots, \lambda_m)$ за реалізацією $x(t)$, $0 \leq t \leq T$ випадкового процесу $\xi(t)$, що є моделлю досліджуваного явища. Кожна з компонент вектора λ_i , $i = \overline{0, m}$ являє собою відповідний інформативний параметр. Реалізацію апіорного вектора E_k для k -го класу буде матриця

$$\mathcal{E}_k = \begin{pmatrix} \lambda_{1,1}^{(k)}, \dots, \lambda_{1,i}^{(k)}, \dots, \lambda_{1,m}^{(k)} \\ \dots \\ \lambda_{j,1}^{(k)}, \dots, \lambda_{j,i}^{(k)}, \dots, \lambda_{j,m}^{(k)} \\ \dots \\ \lambda_{n,1}^{(k)}, \dots, \lambda_{n,i}^{(k)}, \dots, \lambda_{n,m}^{(k)} \end{pmatrix} \quad (1)$$

Кожне рішення щодо класифікації стану об'єкту контролю є результатом статистичних висновків на основі спостережень \mathcal{E} . Припустимо, що \mathcal{E} можна описати багатовимірною щільністю нормального розподілу. Тоді щільність розподілу \mathcal{E}_k у випадку з K класів (гіпотез)

$$P(X) = N(M_k, C_k) = \frac{1}{(2\pi)^{m/2} |C_k|^{1/2}} \exp \left[-\frac{1}{2} (X - M_k)^T C_k^{-1} (X - M_k) \right] \quad (2)$$

де $k = \overline{1, K}$;

кожна щільність розподілу \mathcal{E}_k повністю визначається вектором середніх значень M_k параметрів λ_k та коваріаційною матрицею C_k , що задані як

$$M_k = E_k\{\mathcal{E}\}, \text{ та}$$

$$C_k = E_k\{(X - M_k)(X - M_k)^T\},$$

де $E\{\cdot\}$ позначає оператор математичного сподівання, що визначається на образі k -го класу;

$|C_k|$ – визначник коваріаційної матриці C_k ;

X – аргумент $P(X)$, який при дослідженні конкретних реалізацій приймає їх значення.

При проведенні контролю конкретних виробів математичне сподівання вектора \mathcal{E} приймає конкретні значення, які відповідають конкретному стану зони контролю та можуть слугувати для ідентифікації дефекту. Для прийняття рішення щодо віднесення отриманого \mathcal{E} до одного з K визначених класів (прийняття одєї з K гіпотез) можна використати, наприклад відомий критерій Баеса [3], що мінімізує середній ризик класифікації і полягає в порівнянні відношення правдоподібності з порогом c . Але на практиці врахувати всю показники (ендокринологічні) не є можливим. Тому визначити значення порогів c для класифікації всіх можливих показників не є доцільним. Тому, для практичного застосування, пропонується використовувати альтернативний метод класифікації, що ґрунтується на визначенні відстані між вектором інформативних параметрів досліджуваного об'єкту та векторами, що характеризують задані класи. Перевагою цього методу є також можливість подальшого розбиття існуючих класів на підкласи та проведення процедури класифікації без ускладнення розрахункового алгоритму. З цих міркувань вирішальним буде наступне правило – за результатами контролю приймається така гіпотеза H_k для якої

$$d_k = \min(d(X, \mathcal{E}_k)), \quad (3)$$

де $d(\cdot)$ – прийнята міра відстані між двома багатомірними розподілами;
 $X = \|\lambda_{j,i}\|$, $i = \overline{1, m}$, $j = \overline{1, n}$ – апостеріорна матриця конкретних реалізацій $\xi(t)$, для якої m – кількість інформативних параметрів, n – кількість повторних експериментів по їх визначенню.

Розглянемо міри відстані, які можуть бути застосовані для прийнятої моделі (2). У випадку співпадаючих коваріаційних матриць $C_1 = C_2 = C$ мірою відстані між відповідними розподілами $f_1 = N(M_1, C)$ та $f_2 = N(M_2, C)$ може слугувати відстань Махаланобіса [1]

$$d^2 = (M_2 - M_1)^T (C)^{-1} (M_2 - M_1), \quad (4)$$

яка має наступні властивості:

1. якщо випадкові вектори $Z_i = (X_i, Y_i)$, $i = 1, 2$ відповідні та складаються з двох незалежних векторів X_i, Y_i то $d^2(Z_1, Z_2) = d^2(X_1, X_2) + d^2(Y_1, Y_2)$ – властивість адитивності,
2. якщо $X_i \in N(M_i, C)$, $i = 1, 2, 3$, $|C| \neq 0$, то $d(X_1, X_3) \leq d(X_1, X_2) + d(X_2, X_3)$ – нерівність трикутника.

Методика обрахунку даних за принципом мінімуму відстані:

- 1) розрахунок середнього значення по кожному показнику у кожній групі (основна, дослідна та еталонна група):

$$\bar{X} = \frac{1}{N \sum_{i=1}^N x_i},$$

де N – кількість пацієнтів у відповідній групі,

X_i – поточне значення відповідного показника;

- 2) розрахунок дисперсії по кожному показнику у кожній групі:

$$D = 1/N - 1 \sum_{i=1}^N (X_i - \bar{X})^2;$$

- 3) розрахунок різниці між середніми значеннями основної або дослідної групи з основним значенням здорової групи ($M_2 - M_1$);
 Побудова коваріаційної матриці для кожної групи окремо:

$$C_k = \begin{pmatrix} D_{1,1}^{(k)}, \dots, D_{1,i}^{(k)}, \dots, D_{1,m}^{(k)} \\ \dots \\ D_{j,1}^{(k)}, \dots, D_{j,i}^{(k)}, \dots, D_{j,m}^{(k)} \\ \dots \\ D_{n,1}^{(k)}, \dots, D_{n,i}^{(k)}, \dots, D_{n,m}^{(k)} \end{pmatrix};$$

- 4) перевірка на однорідність дисперсії за критерієм Фішера [2];
- 5) обрахунок відстані Махаланобіса за формулою (4).

Результати обрахунку виявились наступні:

Вибірки є однорідними за критерієм Фішера, отже відстань Махаланобіса є:

- 1) між основною та здоровою групою 34.794,
- 2) між дослідною та здоровою групою 43.87,
- 3) між основною та дослідною групою 1.59. Таке значення говорить про те, що основна та дослідна групи є однорідними.

Після лікування:

- 4) між основною та здоровою групою 28.999,
- 5) між дослідною та здоровою групою 59.412,
- 6) між основною та дослідною групою 3.093.

Як видно з отриманих результатів, відстань між основною групою та групою здорових зменшилась, з цього можна зробити висновок, що лікування було ефективним.

Відстань між дослідною групою та групою здорових збільшилась, з цього можна зробити висновок, що лікування не було ефективним.

Список літератури

1. J.Tou, R.Gonzalez Pattern Recognition Principles Addison– Wesley Publishing Company London 1974. – 378 s.
2. Статистичний аналіз даних вимірювань: навч. посіб. / Єременко В.С., Куц Ю.В., Мокійчук В.М., Самойліченко О.В. – К.: НАУ, 2013.– 320 с.
3. Дорожовець М. Опрацювання результатів вимірювань: навч. посібник / М. Дорожовець. – Львів: Видавництво Національного університету “Львівська політехніка”, 2007. – 624 с