

УДК 519.711:616.12
DOI: 10.15587/1729-4061.2014.31079

МОДЕЛЮВАННЯ КЛАСИФІКАТОРІВ ДИФЕРЕНЦІАЛЬНОЇ ДІАГНОСТИКИ ПАТОЛОГІЧНИХ СТАНІВ СИСТЕМИ КРОВООБІГУ

Є. А. Настенко

Доктор біологічних наук, старший науковий співробітник, завідувачий відділенням Відділ інформаційних технологій та математичного моделювання фізіологічних процесів ДУ «Науково-дослідний інститут серцево-судинної хірургії ім. М. М. Амосова» НАМН України вул. Миколи Амосова, 6, м. Київ, Україна, 03110 E-mail: nastenko@inbox.ru

В. А. Павлов

Кандидат технічних наук, доцент* E-mail: vpavlo@bk.ru

О. К. Носовець

Аспірант, старший викладач* E-mail: e.nosovets@ya.ru

*Кафедра біомедичної кібернетики Національний технічний університет України «Київський політехнічний інститут» пр. Перемоги, 37, м. Київ, Україна, 03056

Наведено результати застосування методів моделювання для створення класифікаторів диференціальної діагностики патологічних станів системи кровообігу для випадку, коли об'єкти характеризуються множинами точок у багатовимірному просторі. Діагностична проблема вирішується шляхом перевodu задачі класифікації у простір ознак-параметрів оптимальної структури об'єктів класифікації. Застосовано методи бінарної логістичної регресії, дискримінантного аналізу та метод групового врахування аргументів

Ключові слова: моделювання, класифікатори, логістична регресія, дискримінантний аналіз, метод групового врахування аргументів

Приведены результаты применения методов моделирования для создания классификаторов дифференциальной диагностики патологических состояний системы кровообращения для случая, когда объекты характеризуются множествами точек в многомерном пространстве. Диагностическая проблема решается путем перевода задачи классификации в пространство признаков-параметров оптимальной структуры объектов классификации. Используются методы бинарной логистической регрессии, дискриминантного анализа и метод группового учета аргументов

Ключевые слова: моделирование, классификаторы, логистическая регрессия, дискриминантный анализ, метод группового учета аргументов

1. Вступ

Провідне місце серед причин смертності населення України займають захворювання системи кровообігу, питома вага яких в структурі загальної летальності складає 66 %, а в працездатному віці – 29,3 % (Державна служба статистики України, 2011 рік) [1].

Для поліпшення даної статистики важливим є створення простого неінвазивного методу діагностики, що дозволить на ранніх етапах виявляти та корегувати патологічні стани системи кровообігу.

Одним з варіантів діагностики станів системи кровообігу, що дозволяє підвищити точність виявлення станів відмінних від нормального, є варіант з застосуванням багаторазових вимірювань показників діастолічного, систолічного артеріального тиску (АТД, АТС) та частоти серцевих скорочень (ЧСС). Це дозволяє підвищити точність класифікації за рахунок підвищення кількості показників, що характеризують кожного пацієнта. Однак, класифікація пацієнтів у в такому випадку не є однозначною.

Це є поширений випадок варіанту задач класифікації, коли об'єкт (пацієнт) характеризується не

поодинокими вимірами в багатовимірному просторі ознак, а їх підмножинами, причому для кожного з них допускається частковий перетин підмножин областей простору ознак де вимірювалися об'єкти різних класів (діагнозів) (рис. 1)

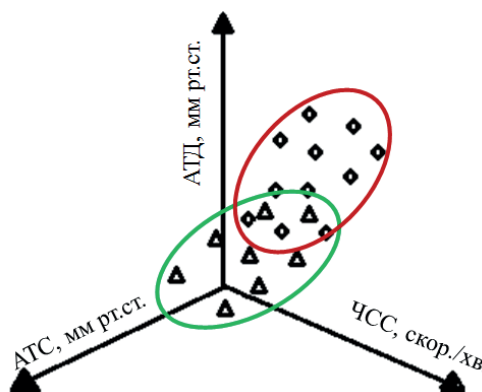


Рис. 1. Приклад перетину показників об'єктів, що належать до різних класів

Така умова не дозволяє при класифікації об'єкта використовувати окремі точки, а вимагає застосування спеціальних підходів, де можливо оперувати множинами вимірювань, або характеристиками, отриманими на основі цих множин, або параметрами різних доцільних розкладань наявних характеристик в модельний ряд [2–4].

2. Аналіз літературних даних та постановка проблеми

Результати досліджень показників артеріального тиску та частоти серцевих скорочень показали, що вони є високоінформативними та відображають функцію серця, функцію ендотелію та стан капілярних судин [5–7]. Ця інформація дала можливість розробити нові підходи до виявлення патологічних станів системи кровообігу [8–10], що базуються на одноразових та багаторазових вимірюваннях показників отриманих у стані спокою. При цьому методи, що не виключають з аналізу дані моніторингу тиску та пульсу під дією психо-фізичного навантаження, майже не розвинуті, в зв'язку з відсутністю норм для таких даних та складністю їх віднесення до певного стану.

Особливістю цієї задачі є неможливість гарантованої, в загальному випадку, класифікації об'єктів по окремих точках Z_i у початковому просторі ознак z_1, \dots, z_M . Такий висновок слідує з можливих перетинів областей простору Q_d , коли об'єкти (пацієнти) d_i і d_j належать різним класам.

Вирішення проблеми пропонується за рахунок переведення задачі класифікації у просторі ознак - параметрів деякої найкращої структури, що описує відомі нам об'єкти класифікації. Така структура одержується комбінаторним алгоритмом МГВА, що базується на основних рішеннях наведених в роботі [11]. Відмінністю запропонованої реалізації [12] від роботи [11] є застосування доцільного для даної задачі зовнішнього критерію: критерію регулярності похибок моделей всіх об'єктів класифікації.

3. Ціль та задачі дослідження

Ціль даної роботи полягає у вирішенні задачі моделювання класифікаторів диференційної діагностики патологічних станів системи кровообігу багатовимірному просторі, що базується на даних отриманих після згортки в точки в просторі параметрів R .

Для досягнення поставленої мети вирішувалися наступні задачі:

- визначити структуру моделі, що реалізує згортку сукупності об'єктів;
- розрахувати класифікатори патологічних станів за різними методами математичного моделювання;
- визначити підхід до математичного моделювання, що має найбільш точні результати класифікації.

4. Клінічні матеріали та методи моделювання класифікаторів диференційної діагностики

В якості дослідницького матеріалу були використані дані багаторазових вимірювань показників АТД,

АТС та ЧСС, отримані в результаті моніторингу у 62 обстежених. Вік обстежуваних знаходився в межах від 33 до 64 років (середній вік 44,6 років). Всього було отримано 12 357 вимірювань за допомогою апарату A&D Medical UA-878 (Японія). 2437 вимірів не брали участь процесі моделювання, оскільки були відібрані для оцінки якості отриманих моделей.

Первинні ознаки пацієнтів – ЧСС, АТС, АТД. В якості вихідної змінної (2) – визначено діастолічний тиск.

В просторі параметрів клінічних ознак x_i , $i=1, \dots, m$ було задано 5 класів (станів системи кровообігу): D_1 – систолічна ізольована гіпертензія (ІГс), D_2 – артеріальна гіпертензія (АГ), D_3 – нормальна регуляція (НР), D_4 – серцева недостатність невисоких ступенів (АНн), D_5 – серцева артеріальна недостатність високих ступенів (АНв). Всі пацієнти, що брали участь у обстеженнях було віднесено до одного з цих класів.

5. Результати створення методу діагностики патологічних станів в багатовимірному просторі

5.1 Реалізація переходу до альтернативного простору ознак

Для вирішення задачі класифікації в багатовимірному просторі припускаємо, що опис об'єкту d задається підмножиною точок $Z^d = \{Z_i\}_d$, знятих з деякої невідомої нам характеристики об'єкту $f_d(z) = 0$. Тоді вказаних характеристик передбачається стільки, скільки об'єктів:

$$f_d(z) = 0, \quad d = 1, \dots, n, \quad (1)$$

Далі, без втрати загальності, розглядаємо той випадок, коли серед початкових змінних z_1, \dots, z_M можливо виділити характерну вихідну змінну y .

Визначимо новий простір ознак x , як простір узагальнених змінних x_1, x_2, \dots, x_{M1} початкового простору z , що найкраще представляють характеристики $f_d(z) = 0$, $d = 1, \dots, n$ по початковій множині точок $Z^d = \{Z_i\}_d$ уже як лінійну згортку по x_i . Тоді з точністю до переозначення, характеристики (1) можливо шукати у вигляді (2):

$$y = r_0 + \sum_{i=1}^M r_i \phi_i(z) = r_0 + \sum_{i=1}^M r_i x_i, \quad (2)$$

де M – розмірність нового простору узагальнених змінних x для представлення об'єктів d .

Тоді рішення задачі класифікації можливо перевести з простору узагальнених змінних, у простір

параметрів $R = \{r_j\}_{j=0, \dots, M}$ характеристик $y = r_0 + \sum_{i=1}^M r_i x_i$,

що дозволить розглядати об'єкти d уже не як множини $Z^d = \{Z_i\}_d$, та не як характеристики $f_d(z) = 0$, $d = 1, \dots, n$, а як точки в просторі параметрів R . Одиночні точки r_d вже однозначно визначають об'єкти d зважаючи на відсутність повністю співпадаючих підмножин $Z^d = \{Z_i\}_d$.

Таким чином вирішення проблеми пропонується за рахунок переведення задачі класифікації у просторі ознак – параметрів деякої найкращої структури, що описує відомі нам об'єкти класифікації. Така струк-

тура одержується комбінаторним алгоритмом МГВА, що базується на основних рішеннях наведених в роботі [4]. Відмінністю запропонованої реалізації [5] від роботи [4] є застосування доцільного для даної задачі зовнішнього критерію: критерію регулярності похибок моделей всіх об'єктів класифікації.

Надалі передбачається класифікувати об'єкти d як точки r_d у просторі параметрів R .

Застосування вказаного алгоритму, дало можливість перевести задачу класифікації у простір параметрів, для чого було обрано оптимальну структуру, одержану за комбінаторним алгоритмом МГВА [5], що описувала вихідну змінну (АТД) за вхідними змінними (АСТ та ЧСС).

$$DIA = r_0 + r_1 \cdot \sqrt[3]{HR} \cdot \frac{1}{SYS} \tag{3}$$

На знайденій структурі було розраховано коефіцієнти моделей для всіх пацієнтів та здійснено перетворення бази даних пацієнтів відповідно до знайдених значень параметрів оптимальної структури (3). Для цього були використані методи нелінійного регресійного аналізу. В подальшому перетворена база даних використана для побудови класифікаторів диференціальної діагностики.

Для побудови діагностичних класифікаторів використовуються наступні методи:

- логістичний регресійний аналіз;
- дискримінантний аналіз;
- метод групового урахування аргументів (МГУА) [6].

В якості остаточних будуть обрані класифікатори, що показали максимальну якість розділення на тестовій вибірці пацієнтів – показники чутливості та специфічності для пацієнтів, що не брали участь в налаштуванні відповідного класифікатора

5. 2. Розрахунок класифікаторів за допомогою дискримінантного аналізу

1. Формула класифікатора, що диференціює стан «ІГс» (табл. 1):

$$Y(x) = -4,672 + 0,074 \cdot r_0 + 0,003 \cdot r_1 \tag{4}$$

Таблиця 1

Характеристики отриманого класифікатора ІГс

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	89,1 %	0,99	0,88
Навчання	90,1 %	1	0,89
Екзамен	84,2 %	0,95	0,83

2. Формула класифікатора, що диференціює стан «АГ» (табл. 2):

$$Y(x) = -3,452 + 0,050 \cdot r_0 + 0,001 \cdot r_1 \tag{5}$$

Таблиця 2

Характеристики отриманого класифікатора АГ

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	68 %	0,71	0,68
Навчання	68,4 %	0,72	0,68
Екзамен	66,2 %	0,68	0,66

3. Формула класифікатора, що диференціює стан «НР» (табл. 3):

$$Y(x) = -3,647 + 0,053 \cdot r_0 + 0,001 \cdot r_1 \tag{6}$$

Таблиця 3

Характеристики отриманого класифікатора НР

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	79 %	0,86	0,76
Навчання	79,1 %	0,86	0,76
Екзамен	78,5 %	0,84	0,76

4. Формула класифікатора, що диференціює стан «СНн» (табл. 4):

$$Y(x) = -0,247 + 0,014 \cdot r_0 + 0,002 \cdot r_1 \tag{7}$$

Таблиця 4

Характеристики отриманого класифікатора СНн

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	70,9 %	0,9	0,66
Навчання	71,1 %	0,905	0,663
Екзамен	69,8 %	0,89	0,65

5. Формула класифікатора, що диференціює стан «СНв» (табл. 5):

$$Y(x) = -5,895 + 0,095 \cdot r_0 + 0,004 \cdot r_1 \tag{8}$$

Таблиця 5

Характеристики отриманого класифікатора СНв

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	91,4 %	0,99	0,859
Навчання	92,1 %	1	0,869
Екзамен	87,9 %	0,98	0,813

Результати моделювання класифікаторів за допомогою дискримінантного аналізу показали великий розкид точності – від 68 % правильно класифікованих при діагностиці стану «АГ», до 91,4 % – при діагностиці стану СНв. Розраховані показники чутливості та специфічності, що є найбільш значимими при оцінці якості моделювання також не показали високих значень.

5. 3 Розрахунок класифікаторів за допомогою логістичної регресії

1. Формула класифікатора, що диференціює стан «ІГс» (табл. 6):

$$Y(x) = 36,445 - 0,406 \cdot r_0 - 0,014 \cdot r_1 \tag{9}$$

Таблиця 6

Характеристики отриманого класифікатора ІГс

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	98 %	0,904	0,99
Навчання	98,2 %	0,909	0,99
Екзамен	97 %	0,88	0,98

2. Формула класифікатора, що диференціює стан «АГ» (табл. 7):

$$Y(x)=4,853-0,032\cdot r_0-0,001\cdot r_1. \quad (10)$$

Таблиця 7

Характеристики отриманого класифікатора АГ

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	90 %	0	0,99
Навчання	90,1 %	0	1
Екзамен	89,3 %	0,1	0,98

3. Формула класифікатора, що диференціює стан «НР» (табл. 8):

$$Y(x)=-4,530-0,045\cdot r_0-0,001\cdot r_1. \quad (11)$$

Таблиця 8

Характеристики отриманого класифікатора НР

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	69,5 %	0,16	0,938
Навчання	70,1 %	0,16	0,942
Екзамен	66,7 %	0,1	0,92

4. Формула класифікатора, що диференціює стан «СНн» (табл. 9):

$$Y(x)=1,579-0,008\cdot r_0-0,001\cdot r_1. \quad (12)$$

Таблиця 9

Характеристики отриманого класифікатора СНн

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	77,5 %	0,03	0,95
Навчання	78,4 %	0,045	0,967
Екзамен	73 %	0	0,91

5. Формула класифікатора, що диференціює стан «СНв» (табл. 10):

$$Y(x)=-28,505+0,521\cdot r_0+0,021\cdot r_1. \quad (13)$$

Таблиця 10

Характеристики отриманого класифікатора СНв

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	95,5 %	0,921	0,976
Навчання	96,3 %	0,928	0,985
Екзамен	91,7 %	0,891	0,934

Моделювання за допомогою методу відтворення бінарної логістичної кривої показало більш точні показники. Найбільша точність класифікації була досягнута для стану «ІГс» – 98 % правильно класифікований значень. Найменшу точність показав класифікатор стану «НР» – 69,5 %. Однак, аналіз показників чутливості показав, що на трьох моделях з п'яти, незважаючи на високі показники правильно класифікований значень, не перевищував значення 0,16.

5.4. Розрахунок класифікаторів за допомогою МГУА

1. Формула класифікатора, що диференціює стан «ІГс» (табл. 11):

$$Y(x)=2,156-0,022\cdot r_0-0,001\cdot a_1-0,01\cdot r_1^3. \quad (14)$$

Таблиця 11

Характеристики отриманого класифікатора ІГс

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	94 %	1	0,93
Навчання	93,7 %	1	0,93
Екзамен	95,4 %	0,99	0,95

2. Формула класифікатора, що диференціює стан «АГ» (табл. 12):

$$Y(x)=1,47-0,013\cdot r_0-0,004\cdot r_1. \quad (15)$$

Таблиця 12

Характеристики отриманого класифікатора АГ

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	87,1 %	0,88	0,87
Навчання	86,4 %	0,88	0,86
Екзамен	90,8 %	0,89	0,91

3. Формула класифікатора, що диференціює стан «НР» (табл. 13):

$$Y(x)=1,028-0,017\cdot r_0-0,001\cdot r_1+0,189\cdot r_0^3+0,001\cdot r_1^3. \quad (16)$$

Таблиця 13

Характеристики отриманого класифікатора НР

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	82,8 %	0,8	0,83
Навчання	82,7 %	0,8	0,839
Екзамен	83,4 %	0,82	0,84

4. Формула класифікатора, що диференціює стан «СНн» (табл. 14):

$$Y(x)=0,436-0,001\cdot r_1. \quad (17)$$

Таблиця 14

Характеристики отриманого класифікатора СНн

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	86 %	0,85	0,86
Навчання	86,2 %	0,863	0,862
Екзамен	84,8 %	0,84	0,85

5. Формула класифікатора, що диференціює стан «СНв» (табл. 15):

$$Y(x)=-0,612+0,027\cdot r_0+0,001\cdot r_1-0,139\cdot r_1^3. \quad (18)$$

Таблиця 15

Характеристики отриманого класифікатора СНв

Вибірка	% розпізнавання	чутливість	специфічність
Вся вибірка	96,9 %	0,99	0,95
Навчання	97,4 %	1	0,957
Екзамен	94,7 %	0,96	0,938

Аналіз отриманих класифікаторів показав, що найбільшу точність показали класифікатори отримані

за допомогою МГУА. Моделі, отримані за допомогою цього методу мали високі показники чутливості, специфічності та найбільший процент правильно класифікованих значень.

6. Висновки

Вирішена задача діагностики патологічних станів системи кровообігу пацієнтів, що характеризуються множинами вимірів за допомогою моделювання класифікаторів, що використовують дані згортки до простору параметрів характеристик об'єктів класифікації. Реалізація модифікованого комбінаторного МГУА, за

рахунок визначення структури, що найліпшим чином описує всі об'єкти включені аналіз, дозволила перейти з багатовимірного простору багаторазових вимірювань до просту ознак характеристик об'єктів. Перехід до такого простору дозволив уникнути проблему перетину областей значень ознак об'єктів, що належать до різних класів. Розрахунок класифікаторів за допомогою трьох методів моделювання (логістичного регресійного аналізу, дискримінантного аналізу, МГУА) показав, що найкращі результати демонструють моделі отримані за допомогою МГУА. Середнє значення правильно класифікованих значень – 89,6 %, процент правильно класифікованих значень всередині своїх класів – 92,3 %.

Література

1. Медико-демографічна ситуація та організація медичної допомоги населенню у 2011 році [Текст] / за ред. О.В. Аніщенко. – Міністерство охорони здоров'я України, Київ, 2012. – 103 с.
2. Забара, С. С. Метод виділення інваріантних ознак сигналів [Текст] / С. С. Забара, Н. Б. Філімонова, К. Х. Зеленский // Доповіді Національної академії наук України. – 2009. – № 2. – С. 49–55.
3. Редько, И. Н. Оценка области глобальной устойчивости уравнения маятникового типа методами теории классификации объектов [Текст] : 4-я конф. / И. Н. Редько // Нелинейные колебания в механических системах ННГУ, 1996. – С. 45–46.
4. Редько, И. Н. Применение методов теории классификации объектов для оценок областей существования установившихся движений [Текст] / И. Н. Редько, В. Д. Шалфеев // Вестник ННГУ, сер. Радиофизика (Нелинейная динамика – синхронизация и хаос). – 1998. – С. 68–72.
5. Cowley, A. W. Long-term control of arterial blood pressure [Text] / A. W. Cowley, Jr. *Physiol* // *Circular Researchers*. – 1992. – Vol. 72, Issue 1. – P. 231–300.
6. Euler, U. S. Observations on the pulmonary arterial blood pressure [Text] / U. S. Euler, G. Liljestrand // *Acta Physiologica Scandinavica*. – 1996. – Vol. 12, Issue 4. – P. 301–320. doi: 10.1111/j.1748-1716.1946.tb00389.x
7. Arthur, C. Arterial pressure regulation [Text] / C. Arthur, M. D. Guyton, G. Thomas // *The American Journal of Medicine*. – 1992. – Vol. 52, Issue 5. – P. 584–594.
8. O'Rourke, M. Arterial stiffness, systolic blood pressure, and logical treatment of arterial hypertension [Text] / M. O'Rourke // *Hypertension*. – 1990. – Vol. 15, Issue 4. – P. 339–347. doi: 10.1161/01.hyp.15.4.339
9. McKenna, M. The ratio of ankle and arm arterial pressure as an independent predictor of mortality [Text] / M. McKenna, S. Wolfson, L. Kuller // *Atherosclerosis*. – 1991. – Vol. 87, Issues 2. – P. 119–128. doi: 10.1016/0021-9150(91)90014-t
10. McVeigh, G. E. Age-Related Abnormalities in Arterial Compliance Identified by Pressure Pulse Contour Analysis [Text] / G. E. McVeigh, C. W. Bratteli, D. J. Morgan // *Hypertension*. – 1999. – Vol. 33, Issue 6. – P. 139–142. doi: 10.1161/01.hyp.33.6.1392
11. Степашко, В. С. Комбинаторный алгоритм МГУА с оптимальной схемой перебора моделей [Текст] / В. С. Степашко // *Автоматика*. – 1981. – № 3. – С. 31–36.
12. Nastenka, Ie. A. Combinatorial algorithm for constructing a parametric feature space for the classification of multidimensional models [Text] / Ie. A. Nastenko, G. V. Knishov, O. K. Nosovets, V. A. Pavlov, N. V. Kondrashova // *Cybernetics and Systems Analysis*. – 2014. – Vol. 50, Issue 4. – P. 627–633. doi: 10.1007/s10559-014-9651-3