

Медицина

УДК 616-006.04

**Павлов Володимир Анатолійович**

*кандидат технічних наук, доцент,*

*доцент кафедри біомедичної кібернетики*

*Національний технічний університет України*

*«Київський політехнічний інститут імені Ігоря Сікорського»*

**Павлов Владимир Анатолиевич**

*кандидат технических наук, доцент,*

*доцент кафедры биомедицинской кибернетики*

*Национальный технический университет Украины*

*«Киевский политехнический институт имени Игоря Сикорского»*

**Pavlov Vladimir**

*Candidate of Technical Sciences, Docent,*

*Associate Professor of the Department of Biomedical Cybernetics*

*National Technical University of Ukraine*

*«Igor Sikorsky Kyiv Polytechnic Institute»*

**Трофименко Олександр Володимирович**

*студент факультету «Біомедичної інженерії»*

*Національного технічного університету України*

*«Київський політехнічний інститут імені Ігоря Сікорського»*

**Трофименко Александр Владимирович**

*студент факультета «Биомедицинской инженерии»*

*Национального технического университета Украины*

*«Киевский политехнический институт имени Игоря Сикорского»*

**Trofymenko Olexandr**

*Student of Biomedical Engineering Faculty of the  
National Technical University of Ukraine  
«Igor Sikorsky Kyiv Polytechnic Institute»*

**Грішко Дмитро Юрійович**

*студент факультету «Біомедичної інженерії»  
Національного технічного університету України  
«Київський політехнічний інститут імені Ігоря Сікорського»*

**Гришко Дмитрий Юрьевич**

*студент факультета «Биомедицинской инженерии»  
Национального технического университета Украины  
«Киевский политехнический институт имени Игоря Сикорского»*

**Hrishko Dmytro**

*Student of Biomedical Engineering Faculty of the  
National Technical University of Ukraine  
«Igor Sikorsky Kyiv Polytechnic Institute»*

**СТРУКТУРНИЙ СИНТЕЗ ЗА КРИТЕРІЄМ РОЗДІЛЬНОСТІ В  
ЗАДАЧІ КЛАСИФІКАЦІЇ ОБ'ЄКТІВ-МНОЖИН  
СТРУКТУРНИЙ СИНТЕЗ ПО КРИТЕРІЮ РАЗРЕШЕНИЯ В  
ЗАДАЧАХ КЛАССИФИКАЦИИ ОБЪЕКТОВ-МНОЖЕСТВ  
STRUCTURAL SYNTHESIS BY CRITERION OF SUSTAINABILITY IN  
THE PROBLEM OF CLASSIFICATION OF OBJECTS-MULTIONS**

*Анотація. У статті розглянуто проблему класифікації множин, де  
об'єкти визначенні, як множини багатовимірних спостережень.*

Вирішення проблеми запропоновано проводити у просторі параметрів моделей об'єктів класифікації що побудовано на структурі яка вибрана з умови найкращого значення критерію роздільності. Для пошуку такої структури розроблено алгоритм методу групового врахування аргументів, зовнішнім критерієм якого вибрано критерій роздільності у трьох формах: внутрішньокласова дисперсія, міжкласова дисперсія та їх відношення.

**Ключові слова:** простір параметрів, метод групового врахування аргументів, оптимальна структура, критерій роздільності.

**Анотація.** В статті розглянута проблема класифікації множин, де об'єкти визначені, як множини багатовимірних спостережень. Розв'язання проблеми пропонується проводити в просторі параметрів моделей об'єктів класифікації побудованих на структурі яка вибрана з умови найкращого значення критерію роздільності. Для пошуку такої структури розроблено алгоритм методу групового врахування аргументів, зовнішнім критерієм якого вибрано критерій роздільності у трьох формах: внутрішньокласова дисперсія, міжкласова дисперсія та їх відношення.

**Ключевые слова:** пространство параметров, метод группового учета аргументов, оптимальная структура, критерий разрешения.

**Summary.** The article deals with the problem of classification of sets, where objects are defined as sets of multidimensional observations. The solution of the problem is proposed to carry out in the space of parameters of models of objects of classification built on the structure which is selected from the condition of the best value of the criterion of separation. To find such a

*structure, an algorithm for the method of group consideration of arguments is developed, the external criterion of which is the criterion of separation in three forms: in-class variance, intercolumn variance and their relation.*

**Key words:** *space of parameters, method of group consideration of arguments, optimal structure, criterion of separation.*

**Постановка задачі.** Важливу роль у функціональності медичних інформаційних систем відіграють роль підсистеми підтримки прийняття рішень. Математичною основою таких систем є найчастіше методи класифікації. Умови роботи системи в цих областях характеризуються високим ступенем невизначеності поведінки суб'єктів завдання і наявністю прихованих змінних, знання яких необхідно для прийняття управлінських рішень.

Відомою проблемою завдань розпізнавання образів є неоднозначність рішення при недостатній інформативності опису об'єкта класифікації. При цьому, як правило, ми маємо справу з одноразовими спостереженнями його характеристик (ознак). Якщо врахувати, що значення ознак об'єкта можуть змінюватися в залежності від значення деякого неконтрольованого параметра, або в цілому від стану середовища то проблема стає ще більш очевидною. Труднощі виникають через можливість часткового перетину областей значень ознак у вихідному просторі вимірюваних змінних для об'єктів з різних класів при різних станах середовища, наслідком чого стає неоднозначність результату класифікації

Певним виходом з положення є опис об'єкта не одним, а множиною спостережень, здійснених при різних умовах. Такий комплекс

спостережень дозволяє більш точно описати об'єкт, як деяку множину його станів у вихідному багатовимірному просторі. Наведемо постановку задачі класифікації об'єктів заданих множинами.

Нехай існує множина класів  $C = \{\bar{C}^1, \bar{C}^2, \dots, \bar{C}^K\}$ , де  $\bar{C}^1 = \{c^1\}$ ,  $\bar{C}^2 = \{c^2\}$ , ...  $\bar{C}^K = \{c^K\}$ , та  $\bar{C}^i \cap \bar{C}^j = \emptyset$  при  $i \neq j$ . На практиці, у базах даних, класи задаються, як апроксимації природніх класів,  $C^K \subset \bar{C}^K$ , де  $C^K = \{c^K\} \subset \{\bar{C}^K\}$ , де об'єкт  $c^K$  заданий, як множина спостережень  $c^K = \{c_1^K, \dots, c_{n_K}^K\}$  та спостереження мають вигляд вектору ознак у вихідному просторі:  $\mathbf{z} = (z_1, \dots, z_M)^T$ .

При вирішенні даної задачі постає дві проблеми. По-перше, досі розроблені підходи, як правило, передбачають однократне вимірювання ознак об'єкта, по-друге, ми допускаємо частковий перетин областей початкового простору ознак для об'єктів класифікації з різних класів.

**Аналіз останніх досліджень та публікацій.** Задача, що представлена вище має конструктивне вирішення у випадку незалежних ознак, що формують простір опису об'єктів [1]. В деяких часткових випадках для залежних ознак було запропоновано вирішення проблеми у роботах [2; 3]. Проте існує потреба для розробки загальних підходів вирішення задачі класифікації об'єктів заданих множинами спостережень.

**Мета дослідження.** Запропонувати підхід до вирішення задачі класифікації об'єктів, заданих множинами спостережень у випадку залежних ознак.

**Вирішення задачі.** В основу пропозиції вирішення поставленої задачі поставимо саме факт залежності ознак, що описують об'єкти класифікації. Запропонуємо будувати моделі для об'єктів класифікації, що

відображають сенс цих залежностей та будемо вирішувати задачу класифікації у параметрів параметрів сконструйованих моделей.

Для цього поставимо задачу пошуку найкращої структури, параметри моделей об'єктів класифікації якої, будемо застосовувати для переводу задачі класифікації у простір параметрів. Будемо шукати такі структури серед підструктур деякого повного виразу найскладнішої структури  $y = a_0 + \sum_{i=1}^M a_i \varphi_i(z) = a_0 + \sum_{i=1}^M a_i x_i$

Структури, що генеруються для пошуку найкращої одної, мають вигляд  $y_1 = a_0 + a_0 x_{i_1}, \dots; y_k = a_0 + a_1 x_{i_1} + \dots + a_k x_{i_k}, \dots$  і нарощуються послідовно все більшої складності, шляхом додавання на кожному  $k$ -тому етапі найкращої, з точки зору зовнішнього критерію (критерію роздільності), узагальненої змінної (УЗ)  $x_{i_k}$ . Надалі пропонується класифікувати об'єкти, як точки в просторі параметрів цих моделей розмірністю  $M$ . Для пошуку оптимальної структури для відображення об'єктів-множин застосовано алгоритм, розроблений за принципами методу групового урахування аргументів МГУА[4,5]. Такий підхід включає в себе дві важливі переваги:

- моделювання регресій для відображення множин спостережень дозволяє відібрати найбільш інформативні ознаки з вихідного простору та відсіяти ті, що не впливають на класифікацію.
- відображення об'єктів у багатовимірні точки простору параметрів моделей дозволяє застосувати для даної задачі стандартні алгоритми класифікації.

Узагальненні змінні (УЗ), що є складовою доданків, з яких будуються структури, представляють собою комбінації перемножень та кореневих розширень ознак вихідного простору. Кореневі розширення

початкових ознак застосовуються для підвищення ефективності кінцевих структур оптимальної складності.

При нарощуванні структур, УЗ генеруються за допомогою мультиплікативного дерева. Гілки дерева відповідають створеним раніше УЗ. На кожному ряду алгоритму реалізується рух по вузлах дерева зліва направо і зверху вниз. Це дерево перебору дозволяє скорочувати час роботи алгоритму за рахунок відсікання непродуктивного нарощування складності структур, що не дають потрібний приріст в значенні зовнішнього критерію.

Параметри моделей, в які відображаються об'єкти-множини, вираховуються за методом найменших квадратів. Розроблений алгоритм МГУА реалізує перебір вкладеними структурами, таким чином, з метою прискорення розрахунку параметрів, застосована рекурентна версія МНК. Переваги у швидкості рекурентного обчислення досягаються завдяки тому, що наступні значення вектору параметрів обчислюється на основі розрахунків попередніх етапів.

В даній роботі для відбору оптимальної структури запропоновано використання зовнішнього критерію роздільності. Даний критерій ґрунтується на обчисленні дисперсії векторів параметрів моделей, що представляють об'єкти-множини, та спрямований на «розділення» класів у просторі параметрів моделей. Для відбору структур, в якості зовнішнього критерію, застосовується три різновиди критерію роздільності: внутрішньокласовий, міжкласовий та комбінований.

*Внутрішньокласовий критерій.*

Даний критерій відбирає структури, що найщільніше «грубують» об'єкти у просторі параметрів своїх моделей до свого центру класу. Нижче наведена форма критерію.

$$\frac{1}{K} \sum_{j=1}^K \frac{1}{n_j} \sum_{r_i \in R^m} \|r_i - \bar{r}_j\|^2$$

де  $r_i, r_j$  – вектори параметрів,  $K$  – кількість класів

*Міжкласовий критерій.*

Даний критерій відбирає структури, що «віддаляють» центри класів об'єктів один від одного у просторі параметрів своїх моделей. Нижче наведена форма критерію.

$$\frac{1}{\frac{1}{C_K^2} \sum_{i=1}^K \sum_{j=i+1}^K \|\bar{r}_i - \bar{r}_j\|^2}$$

Де  $\bar{r}$  – центри класів,  $K$  – кількість класів

*Комбінований критерій.*

Даний критерій у пропонованій формі успадковує вимоги обох попередніх критеріїв.

$$\frac{\frac{1}{K} \sum_{j=1}^K \frac{1}{n_j} \sum_{r_i \in R^m} \|r_i - \bar{r}_j\|^2}{\frac{1}{C_K^2} \sum_{i=1}^K \sum_{j=i+1}^K \|\bar{r}_i - \bar{r}_j\|^2}$$

**Приклад розрахунку структури та результати класифікації стану серцево-судинної системи.**

Завдання було викликане реальною проблемою класифікації функціональних станів серцево-судинної системи на основі вимірів пульсу і тиску людини в різних обставинах. Наведено групу з 180 пацієнтів, кожен з яких відноситься до одного з п'яти функціональних класів станів серцево-судинної системи:

1. Артеріальна гіпертонія.
2. Ізольована систолічна гіпертонія.
3. Нормальне регулювання.
4. Серцева недостатність низького рівня.
5. Серцева недостатність високого рівня.

Для кожного пацієнта дано набори тривимірних замірів діастолічного тиску, систолічного тиску, частоти серцевих скорочень (DIA, SYS, HR) що вимірювались в кількості від 4 до 151 разів та отримані в різних умовах.

Метою задачі є побудова класифікаторів функціональних станів серцево-судинної системи.

За розробленим алгоритмом та за критерієм роздільності у формі (а саме міжкласової дисперсії) було знайдено оптимальну структуру вигляду:

$$a_0 + a_1 * SYS + a_2 * SYS^2 + a_2 * DIA^2 + a_3 * SYS^2 + a_4 * SYS$$

Для цієї структури було обраховано моделі кожного з 180- пацієнтів, параметри яких стали новими ознаками в задачі класифікації функціональних станів ССС пацієнтів.

Наступним кроком на знайденому наборі ознак застосовано алгоритм класифікації «Випадковий Ліс» розроблений на основі дерев прийняття рішень [6]. Алгоритм було навчено на відображеннях об'єктів у новому просторі параметрів та перевірено на тестовій вибірці для отримання основних метрик класифікації. Тренувальна вибірка складає 135 об'єктів, тестова 45. Результати наведено у таблиці 1.

Таблиця 1

**Значення метрик по класах**

Клас	Точність (%)	Повнота (%)	F1-міра (%)
1	97	76	85
2	61	93	74
3	82	73	77
4	99	62	86
5	99	81	90

**Висновки.** Запропоновано підхід до розпізнавання об'єктів, заданих підмножинами рядків матриці об'єкт-властивості. Підхід передбачає переведення задачі розпізнавання у простір параметрів найкращої структури моделей об'єктів класифікації де вже кожний об'єкт представлений однією багатовимірною точкою в просторі параметрів своєї моделі.

Для знаходженні такої структури розроблено версію рекурентного багатоетапного алгоритму на основі МГУА. Для пошуку оптимальної структури моделей застосовано критерій роздільності, та розглянуто три його різновиди.

Розглянуто приклад для вирішення задачі класифікації функціональних станів ССС людини. Після переведення задачі у простір параметрів моделей об'єктів застосовано алгоритм класифікації «Випадковий ліс». Одержано високі показники якості класифікації.

**Література**

1. Fukunaga, K. (1990). Introduction to statistical pattern recognition (p. 491). Boston: Academic Press.

2. Jung, S., & Qiao, X. (2014). A statistical approach to set classification by feature selection with applications to classification of histopathology images. *Biometrics*, 70(3). P. 536-545. URL: <http://dx.doi.org/10.1111/biom.12164>
3. Войтикова М.В. Хурса Р.В. Линейная регрессия параметров артериального давления для определения риска развития вторичной гипотензии / Артериальная гипертензия. 6(44). 2015. С.38-42
4. Ивахненко А.Г. Помехоустойчивость моделирования[Текст] / А.Г. Ивахненко, В.С. Степашко // Киев: «Наук.думка». 1985. 216 с.
5. Ye. Nastenka, O. Konoval, O. Nosovets, V. Pavlov. Set Classification, pp. 44-83 - In: *Techno-Social Systems for Modern Economical and Governmental Infrastructures (Advances in Finance, Accounting, and Economics)*. : IGI Global; 1 ed. (July 13, 2018), 351 p. DOI: 10.4018/978-1-5225-5586-5.ch003
6. Tin Kam Ho. *Random Decision Forests* (2011).