

**MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE  
SUMY STATE UNIVERSITY  
UKRAINIAN FEDERATION OF INFORMATICS**

## **PROCEEDINGS**

**OF THE IV INTERNATIONAL SCIENTIFIC  
CONFERENCE**

**ADVANCED INFORMATION  
SYSTEMS AND TECHNOLOGIES**

**AIST-2016**



**May 25 –27, 2016  
Sumy, Ukraine**

# Informational Extreme Cluster Analysis of Input Data

T.M. Yefimenko, O.V. Korobchenko  
Sumy State University, Ukraine, elena9191@gmail.com

**Abstract.** *The categorical model and decision support system learning algorithm are considered in the article. Proposed algorithm allows to create decision support system, which is functioning in a cluster-analysis state. Synthesis of the decision support system is based on maximization of informational system ability due to making additional information restrictions in the learning process.*

**Keywords.** *Information-Extreme Intelligent Technology, Cluster Analysis, Training, Optimization, Decision Support System, K-Means Clustering, Decision Rules.*

## ВСТУП

Переважна більшість методів кластер-аналізу застосовує процедуру  $K$ -середніх. Це не дозволяє побудувати чітке розбиття простору ознак на класи розпізнавання через використання дистанційних критеріїв близькості, оскільки апріорно на практиці воно є нечітким.

Пропонується в рамках інформаційно-екстремальної інтелектуальної технології (ІЕІ-технології), яка ґрунтується на максимізації в процесі машинного навчання інформаційної спроможності системи розпізнавання, здійснювати кластер-аналіз вхідних даних з метою побудови класифікованої нечіткої навчальної матриці, на основі якої будуються в процесі інформаційно-екстремального навчання безпомилкові вирішальні правила [1]. Такий підхід є виправданим в задачах розпізнавання образів за наявності словника ознак або алфавіту класів розпізнавання великої потужності, коли формування апріорно класифікованої нечіткої навчальної матриці вимагає від розробника інформаційного забезпечення системи розпізнавання довготривалої рутинної роботи.

Особливо актуальною є автоматизація формування вхідних навчальних матриць із застосуванням кластер-аналізу вхідних даних при розпізнаванні зображень, наприклад, при інформаційному синтезі систем діагностування патологій в медицині, розпізнаванні об'єктів на місцевості тощо.

## АЛГОРИТМ НАВЧАННЯ СИСТЕМИ РОЗПІЗНАВАННЯ

Сформована за результатами кластер-аналізу вхідна класифікована нечітка навчальна матриця в рамках ІЕІ-технології трансформується в бінарну навчальну матрицю, на основі якої в процесі оптимізації параметрів інформаційно-екстремального навчання будуються безпомилкові вирішальні правила. Застосування безпомилкових за навчальною матрицею вирішальних правил при функціонуванні системи розпізнавання в режимі екзамену, тобто безпосереднього розпізнавання, дозволяє отримати повну ймовірність правильного прийняття рішень, наближену до максимальної граничної.

Вхідний математичний опис системи розпізнавання розглянемо у вигляді структури

$\Delta_B = \langle G, T, \Omega, Z, Y', Y, X; \Phi_1, \theta_1, \Phi_2, \Phi_3 \rangle$ ,  
де  $G$  – множина вхідних факторів;  $T$  – множина моментів зчитування інформації;  $\Omega$  – простір ознак розпізнавання;  $Z$  – простір функціональних станів;  $Y'$  – некласифікована навчальна матриця;  $Y$  – класифікована вхідна навчальна матриця;  $X$  – бінарна навчальна матриця;  $\Phi_1, \Phi_2, \Phi_3$  – оператори формування вхідних навчальних матриць;  $\theta_1$  – оператор кластер-аналізу

вхідних даних, який будує за методом К-середніх розбиття  $\mathfrak{R}'$ .

#### КАТЕГОРІЙНА МОДЕЛЬ ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНОГО КЛАСТЕР-АНАЛІЗУ

На рис. 1 показано категорійну модель інформаційно-екстремального навчання, яка є одночасно узагальненою структурною схемою алгоритму навчання системи розпізнавання.

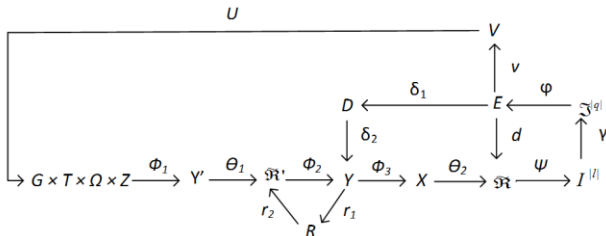


Рисунок 1 – Категорійна модель інформаційно-екстремального кластер-аналізу

Тривалість кластер-аналізу вхідних даних залежить від виконання умови наповнення контейнерів класів розпізнавання заданою мінімальною кількістю реалізацій. З цією метою оператор  $r_1$  обчислює кількість реалізацій в кожному кластері і при невиконанні умови зупину вибирає із множини  $R$  відповідне значення радіуса, а оператор  $r_2$  змінює радіуси кластерів.

В процесі навчання системи розпізнавання оператор  $\theta_2$  будує розбиття  $\mathfrak{R}$ , геометричні параметри якого формують вирішальні правила. Де оператор класифікації  $\psi$  перевіряє основну статистичну гіпотезу про належність реалізації відповідному класу розпізнавання і формує множину гіпотез  $I^{|l|}$ , де  $l$  – потужність множини. Оператор  $\gamma$  обчислює множину точнісних характеристик  $\mathfrak{Z}^{[q]}$ , де  $q = l^2$ . Оператор  $\phi$  обчислює значення інформаційного критерію  $E$ , який є функціоналом точнісних характеристик.

Правий контур оптимізації замикається оператором  $d$ , який змінює геометричні параметри контейнерів класів розпізнавання. Контур оптимізації контрольних допусків на

ознаки розпізнавання замикаються через множину  $D$  – система контрольних допусків.

У рамках ІЕІ – технології вирішальні правила будується спочатку для гіперсферичних контейнерів класів розпізнавання. Якщо в процесі навчання не вдалося побудувати безпомилкові за навчальною матрицею вирішальні правила, то необхідно перейти за допомогою оператора  $v$  до контейнерів більш складної геометричної форми (гіперциліндродної, еліпсоїдної або іншої форми). На рис. 1  $V$  – множина типів вирішальних правил. Оператор  $U$  – регламентує процес інформаційно-екстремального кластер-аналізу.

#### ВИСНОВКИ

Таким чином, використання в рамках інформаційно-екстремального кластер-аналізу дистанційних та інформаційних мір близькості дозволяє сформувати вхідну нечітку класифіковану навчальну матрицю. Реалізація алгоритму навчання у рамках ІЕІ-технології з використанням сформованої навчальної матриці дозволяє трансформувати апріорно нечітке розбиття простору ознак на класи розпізнавання в чітке розбиття, тобто побудувати безпомилкові за навчальною матрицею вирішальні правила і автоматизувати процес формування вхідного математичного опису системи розпізнавання.

#### REFERENCES

- [1] Dovbysh A.S. Fundamentals of intelligent systems: manual. guidances. – Sumy: A publ. is Sumy State University, 2009. - 171.
- [2] Dovbysh A. S. Information-extreme algorithm for recognizing current distribution maps in magnetocardiography / A. S. Dovbysh, S.S. Martynenko, A.S. Kovalenko, N.N. Budnyk // Journal of Automation and Information Sciences. – 2011.– V. 43.– № 2.–P. 63-70.