

# НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

## NEUROINFORMATICS AND INTELLIGENT SYSTEMS

## НЕЙРОИНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК 004.8:004.032.26

### ОНЛАЙН МЕТОД МОЖЛИВІСНОЇ КЛАСТЕРИЗАЦІЇ ДАНИХ НА ОСНОВІ ЕВОЛЮЦІЙНОЇ ОПТИМІЗАЦІЇ КОТЯЧИХ ЗГРАЙ

**Бодяньський С. В.** – д-р техн. наук, професор, професор кафедри штучного інтелекту, Харківський національний університет радіоелектроніки, Харків, Україна.

**Шафроненко А. Ю.** – канд. техн. наук, доцент, доцент кафедри інформатики, Харківський національний університет радіоелектроніки, Харків, Україна.

**Клімова І. М.** – асистент кафедри системної інженерії Харківського національного університету радіоелектроніки, Харків, Україна.

#### АНОТАЦІЯ

**Актуальність.** Проблема кластеризації великих даних сьогодні є дуже цікавою сферою штучного інтелекту. Це завдання часто зустрічається у багатьох додатках, пов'язаних з інтелектуальним аналізом даних, глибоким навчанням, веб-майнінгом тощо. Для вирішення цих проблем традиційні підходи та методи вимагають, щоб уся вибірка даних подавалася у пакетній формі.

**Мета.** Метою роботи було запропонувати метод нечіткої можливісної кластеризації даних з використанням еволюційної оптимізації котячих зграй, який був би позбавлений недоліків традиційних підходів до кластеризації даних.

**Метод.** Процедура нечіткої можливісної кластеризації даних із використанням еволюційних алгоритмів, для більш швидкого визначення екстремумів вибірки, центроїдів кластерів та адаптивних функцій, що дозволяють не витратити ресурси машини для зберігання проміжних розрахунків та не потребують додаткового часу для вирішення поставленої задачі кластеризації даних незалежно від розмірності та способу подачі на обробку.

**Результати.** Запропонований алгоритм кластеризації даних на основі еволюційної оптимізації є достатньо простим у чисельній реалізації, позбавлений недоліків, властивих традиційним методам нечіткої кластеризації та може працювати з великим розміром вихідної інформації, що обробляється, в режимі онлайн у реальному часі.

**Висновки.** Результати експерименту дозволяють рекомендувати розроблений метод для вирішення проблем автоматичної кластеризації та класифікації великих даних, максимально швидко знаходити екстремуми цільової функції, незалежно від способу подачі даних на обробку. Запропонований метод онлайн можливісної нечіткої кластеризації даних на основі еволюційної оптимізації котячих зграй призначений для використання у гібридних системах обчислювального інтелекту, нейро-фаззі системах, в навчанні штучних нейронних мереж, у завданнях кластеризації та класифікації.

**КЛЮЧОВІ СЛОВА:** нечітка кластеризація, правило навчання, оптимізація котячої зграї, режим трасування, режим пошуку.

#### АБРЕВІАТУРИ

FCM – метод нечітких с-середніх;

CSO – алгоритм оптимізації зграї котів.

#### НОМЕНКЛАТУРА

$X$  – матриця набору даних;

$k$  – номер вектору-спостереження;

$i$  – номер атрибуту вектору-спостереження;

$j$  – номер неперетинного класу;

$x(k)$  – вектор-спостереження;

$x_i(k)$  – значення попередньо обробленого вихідного  $k$ -го вектора-спостереження для  $i$ -го атрибута;

$l, q$  – номер кластеру;

$m$  – кількість неперетинних класів;

$U_j(k)$  – рівень належності  $j$ -го кластеру;

$Cl$  – кластер;

$D$  – Евклідова відстань;

$E$  – цільова функція;

$\mu$  – скалярний параметр, який визначає відстань, на якій рівень належності дорівнює 0,5;

$c$  – центроїд кластера;

$c_p$  – режим  $p$ -ої кішки;  
 $\tau$  – ітерація пошуку;  
 $\beta$  – фаззифікатор;  
 $\alpha$  – параметр, який визначає властивості інерції режиму трасування;  
 $\eta$  – параметр крока пошуку;  
 $\Xi(\tau)$  – випадкова складова, яка вносить додаткові стохастичні рухи у процес трасування;  
 $\eta_\xi$  – параметр, який визначає амплітуду рухів.

## ВСТУП

В цей час методи обчислювального інтелекту широко використовуються для вирішення багатьох складних задач аналізу даних. Однією з основних областей обчислювального інтелекту є, так звані, еволюційні алгоритми, які за суттю представляють певні математичні моделі розвитку біологічних організмів.

Проблема аналізу даних, що пов'язана з кластеризацією векторних образів, часто зустрічається у багатьох додатках для інтелектуального аналізу, але останнім часом основна увага приділяється нечіткій кластеризації [1–3] при обробці векторних образів з різним рівнем ймовірності, можливості або належності можуть належати більше ніж одному класу.

Самоорганізовані мапи Кохонена [4] є дуже ефективним засобом узгодженої онлайн-кластеризації, використання якої дозволяє вирішувати завдання в режимі онлайн. Еволюційні алгоритми у цьому випадку дозволяють покращити результати кластеризації даних коли ці дані надходять послідовно у онлайн-режимі.

**Об'єкт дослідження** можливісна кластеризація даних в режимі онлайн.

**Предмет дослідження** процедура онлайн можливоїсної кластеризація даних на основі еволюційної оптимізації котячих зграй.

**Мета роботи** полягає у запровадженні можливоїсної процедури нечіткої кластеризації даних, яка може працювати онлайн, та швидко знаходити екстемуми (центри) кластерів, незалежно від кількості даних, що надходять.

## 1 ПОСТАНОВКА ЗАВДАННЯ

Проблема нечіткої кластеризації масивів даних розглядається в умовах, коли сформовані кластери довільно перетинаються в просторі ознак. Вихідною інформацією для вирішення проблеми є масив багатовимірних векторів даних, сформований на вибірці спостережень  $X = \{x(1), \dots, x(k), \dots, x(N)\} \subset R^n$ , де  $k$  – у загальному випадку номер спостереження у вихідному масиві,  $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T \in R^n$ . Результатом кластеризації є розділ цього масиву на  $m$  неперетинних класів  $Cl_j$  з прототипами-

центроїдами  $Cl_j \in R^n$ ,  $j = 1, 2, \dots, m$  та обчислення рівнів належності  $0 \leq U_j(k) \leq 1$  кожного спостереження  $x(k)$  до кожного кластера  $Cl_j$ .

## 2 ОГЛЯД ЛІТЕРАТУРИ

Основний недолік класичних імовірнісних алгоритмів кластеризації [1–3] пов'язаний з обмеженнями рівнів належності, сума яких повинна дорівнювати одиниці. Така ситуація призвела до створення можливоїсних нечітких алгоритмів кластеризації [5, 6].

Цільова функція можливоїсної кластеризації має вигляд

$$E(U_j(k), c_j, \mu_j) = \sum_{k=1}^N \sum_{j=1}^m U_j^\beta(k) D^2(\tilde{x}_k, c_j) + \sum_{j=1}^m \mu_j \sum_{k=1}^N (1 - U_j(k))^\beta. \quad (1)$$

Мінімізація функції (1) за  $U_j(k)$ ,  $c_j$  та  $\mu_j$  веде до системи рівнянь

$$\begin{cases} U_j^{(\tau+1)}(k) = \frac{1}{1 + \left(\frac{D^2(\tilde{x}(k), c_j^{(\tau)})}{\mu_j^{(\tau)}}\right)^{\frac{1}{\beta-1}}}, \\ c_j^{(\tau+1)} = \frac{\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta \tilde{x}(k)}{\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta}, \\ \mu_j^{(\tau+1)} = \frac{\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta D^2(\tilde{x}(k), c_j^{(\tau+1)})}{\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta}, \end{cases} \quad (2)$$

у випадку, коли  $\beta = 2$  ми приходимо до рішення, яке має вигляд (аналог класичного FCM [1]):

$$\begin{cases} U_j^{(\tau+1)}(k) = \frac{1}{1 + \frac{\|\tilde{x}(k) - c_j^{(\tau)}\|^2}{\mu_j^{(\tau)}}}, \\ c_j^{(\tau+1)} = \frac{\sum_{k=1}^N (U_j^{(\tau)}(k))^2 \tilde{x}(k)}{\sum_{k=1}^N (U_j^{(\tau)}(k))^2}, \\ \mu_j^{(\tau+1)} = \frac{\sum_{k=1}^N (U_j^{(\tau)}(k))^2 \|\tilde{x}(k) - c_j^{(\tau+1)}\|^2}{\sum_{k=1}^N (U_j^{(\tau)}(k))^2}. \end{cases} \quad (3)$$

В режимі онлайн формули (2), (3) можна записати у вигляді [6, 7]

$$\left\{ \begin{aligned} U_j(k+1) &= \frac{1}{1 + \left( \frac{D^2(\tilde{x}(k+1), c_j(k))}{\mu_j(k)} \right)^{\frac{1}{\beta-1}}}, \\ c_j(k+1) &= c_j(k) + \eta(k+1)U_j^\beta(k+1)(\tilde{x}(k+1) - c_j(k)), \\ \mu_j(k+1) &= \frac{\sum_{p=1}^{k+1} U_j^\beta(p) D^2(\tilde{x}(p), c_j(k+1))}{\sum_{p=1}^{k+1} U_j^\beta(p)} \end{aligned} \right. \quad (4)$$

або

$$\left\{ \begin{aligned} U_j(k+1) &= \frac{1}{1 + \frac{\|\tilde{x}(k) - c_j(k)\|^2}{\mu_j(k)}}, \\ c_j(k+1) &= c_j(k) + \eta(k+1)U_j^2(k+1)(\tilde{x}(k+1) - c_j(k)), \\ \mu_j(k+1) &= \frac{\sum_{p=1}^{k+1} U_j^2(p) \|\tilde{x}(p) - c_j(k+1)\|^2}{\sum_{p=1}^k U_j^2(p)}, \end{aligned} \right. \quad (5)$$

що дозволяє вирішувати нечітку задачу кластеризації в режимі онлайн.

Для пошуку глобального екстремуму (1) доцільно використовувати, так звані, алгоритми оптимізації рою частинок, а саме еволюційних частинок [8], що розробляються та досліджуються в цей час в рамках загальної теорії та практики обчислювального інтелекту. Серед алгоритмів роїння одним з найшвидших є так звані алгоритми зграї котів [9–12], які виявилися ефективними у вирішенні широкого кола завдань з обробки даних.

### 3 МАТЕРІАЛИ І МЕТОДИ

Оптимізація котячої зграї передбачає, що кожна кішка зграї може перебувати в одному з двох станів: режимі пошуку та режимі трасування (відстеження). У цьому випадку режим пошуку пов'язаний з повільними рухами з невеликою амплітудою навколо початкової позиції (сканування простору в районі поточної позиції) та режимом трасування, який визначається швидкими стрибками з великою амплітудою і дозволяє кожній конкретній кішці вискочити з локального екстремуму, якщо вона там є.

Поєднання локального сканування та швидких змін поточного стану дозволяє збільшити ймовірність знаходження глобального екстремуму порівняно з традиційними методами багатоекстремальної оптимізації.

У загальному випадку обидва ці режими для кожного із зграї котів можуть бути описані рекурентною процедурою оптимізації

$$c_p(\tau+1) = c_p(\tau) - \alpha(c_p(\tau) - c_p(\tau-1)) - \eta \hat{\nabla} E_M(c_p(\tau)) + \eta_\xi \Xi(\tau). \quad (6)$$

Так, коли  $\alpha = 0$ , приходимо до стандартного градієнтного пошуку, тобто режиму пошуку у зграї, при  $0 < \alpha < 1$  процес оптимізації набуває інерційних властивостей типу «важкої кулі», але він може не зупинятися поблизу глибокого екстремуму.

Таким чином, кожна кішка може одночасно знаходитись у режимі пошуку та відстеження та, маючи достатню кількість котів у зграї, можна забезпечити пошук глобального екстремуму.

### 4 ЕКСПЕРИМЕНТИ

Онлайн алгоритм можливої нечіткої кластеризації на основі еволюційної оптимізації котячого рою був реалізований на двох масивах даних: табл. 1. Порівняльний аналіз кластеризації вихідних даних був проведений за допомогою нечітких С-середніх, адаптивного можливої нечіткого методу, метода Гауфсона-Кесселя та можливої нечіткої кластеризації на основі еволюційної оптимізації котячої зграї. Параметри моделювання представлені у табл. 2.

Таблиця 1 – Зразки даних

Data set	Data Number	Attributes Number	Cluster Number	Data Source
Gas	296	2	6	Box and Jenkins (1970)
Glass	214	9	6	Mackey and Glass (1977)

Таблиця 2 – Параметри алгоритму CSO

Parameters	Value
SRD	Random [0,1]
Seeking memory Pool (SMP)	5
Population size	Number of clusters
r1	Random in [0,1]
c1	Const
SPC	Random in [0,1]
Number of iteration	Manually

### 5 РЕЗУЛЬТАТИ

Було проведено порівняльний аналіз якості даних кластеризації за основними характеристиками рейтингів якості таких, як швидкість кластеризації даних та середня похибка.

В табл. 3 та табл. 4 наведені результати роботи алгоритмів, запропонованих для порівняння з різною кількістю спостережень.

Таблиця 3 – Порівняльні характеристики середньої похибки для вибірки Gas з різною кількістю спостережень в відсотках

Метод	50	Час	100	Час	150	Час
FCM	1,62	1,19	1,35	2,55	0,98	3,03
GK	1,66	1,62	1,32	2,72	0,99	3,12
Адаптивний можливої нечіткої кластеризації	1,22	1,15	1,02	2,02	0,75	2,10
Онлайн алгоритм можливої нечіткої кластеризації даних на основі еволюційної оптимізації котячих зграї	0,69	1,02	0,49	1,33	0,14	1,41

Таблиця 4 – Порівняльні характеристики середньої похибки для вибірки Glass з різною кількістю спостережень у відсотках

Метод	50	Час	100	Час	150	Час
FCM	1,74	1,21	1,44	2,40	0,86	3,10
GK	1,85	1,74	1,53	2,82	0,99	3,27
Адаптивний можливісний метод нечіткої кластеризації	1,43	1,36	1,22	2,55	0,65	2,60
Онлайн алгоритм можливісної класте- ризації даних на основі еволюційної оптимізації котячих зграй	1,11	1,17	1,00	1,23	0,54	1,11

Аналізуючи отримані результати, можна зробити висновок, що незалежно від розміру вихідної інформації, що подається на обробку запропонованим методом для порівняння працездатності та ефективності, запропонований підхід до можливісної кластеризації даних на основі еволюційного методу котячих зграй не поступається швидкодією та якістю кластерування у порівнянні з відомими алгоритмами.

Порівняльний аналіз продемонстровано на діаграмах залежності похибки та часу від кількості спостережень для різних вибірок даних.

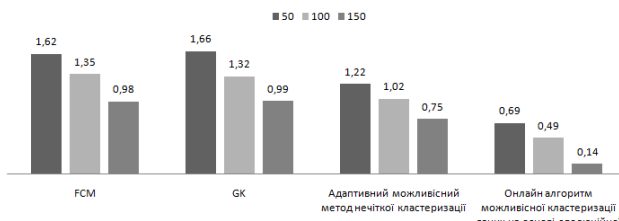


Рисунок 1 – Діаграма залежності похибки від кількості спостережень (50, 100, 150) для вибірки Gas



Рисунок 2 – Діаграма залежності похибки від кількості спостережень (50, 100, 150) для вибірки Glass

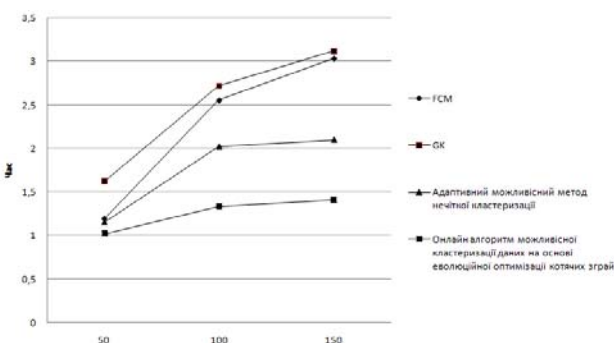


Рисунок 3 – Діаграма залежності часу кластеризації від кількості спостережень (50, 100, 150) для вибірки Gas

© Бодяньський С. В., Шафроненко А. Ю., Клімова І. М., 2021  
DOI 10.15588/1607-3274-2021-2-7

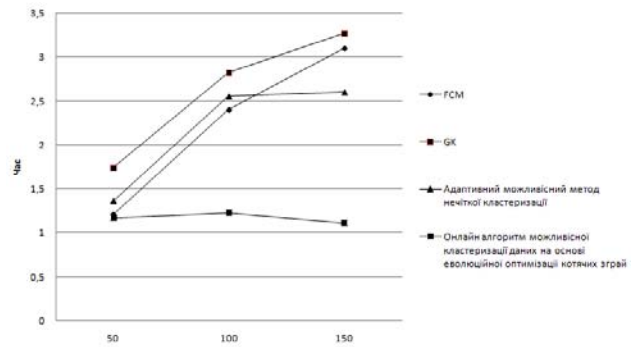


Рисунок 4 – Діаграма залежності часу кластеризації від кількості спостережень (50, 100, 150) для вибірки Glass

Як видно із діаграм, наведених на рисунках, швидкість та похибка в запропонованому методі можливісної кластеризації даних на основі еволюційного підходу демонструє достатньо високі показники.

## 6 ОБГОВОРЕННЯ

Результати кластеризації наборів даних, наведених в таблицях, демонструють швидкодією та якістю роботи методів кластеризації. Запропонований онлайн алгоритм можливісної кластеризації даних на основі еволюційної оптимізації котячих зграй демонструє гарні результати роботи.

Порівняльний аналіз запропонованого алгоритма проводився з відомими на сьогодні алгоритмами кластеризації такими, як FCM, Густафсон-Кессель-алгоритм і адаптивний можливісний метод нечіткої кластеризації.

На рисунках, що наведені вище, продемонстрована робота алгоритмів у порівнянні із запропонованим онлайн алгоритмом можливісної кластеризації даних на основі еволюційної оптимізації котячих зграй.

Завдяки своїй адаптивності та функціям еволюційної оптимізації котячих зграй, алгоритм не потребує багато часу для обробки даних, що надходять у реальному часі, та не завантажує себе проміжними розрахунками за рахунок функцій адаптивності. Це досить яскраво демонструють діаграми залежності часу кластеризації від кількості спостережень та залежності похибки від кількості спостережень.

## ВИСНОВКИ

Запропонований підхід є досить простим в чисельній реалізації, має високу продуктивність і забезпечує високу якість нечіткої кластеризації великих масивів. Показано, що задачу можна вирішити на основі самоорганізованої нейро-нечіткої мережі Кохонена «Переможець отримує більше» для нечіткої кластеризації та алгоритмів оптимізації котячої зграї.

**Наукова новизна:** вперше запропонована можливісна нечітка кластеризація даних, яка може працювати онлайн, та швидко знаходити екстемуми (центри) кластерів, не залежно від кількості даних, що надходять.

**Практичне значення:** результати експерименту дозволяють рекомендувати запропоновані методи для використання на практиці для вирішення проблем автоматичної кластеризації великих даних.

**Перспективи подальших досліджень** методи нечіткої кластеризації даних для широкого класу практичних проблем.

### ПОДЯКА

Робота виконана в рамках науково-дослідного проекту державного бюджету Харківського національного університету радіоелектроніки «Глибокі гібридні системи обчислювального інтелекту для аналізу потоків даних та їх швидке навчання» (номер державної реєстрації 0119U001403).

### ЛІТЕРАТУРА / LITERATURA

1. Bezdek J. C. Pattern recognition with fuzzy objective function algorithms / J. C. Bezdek. – New York : Springer, 1981. – 253p. DOI <https://doi.org/10.1007/978-1-4757-0450-1>.
2. Höppner F. Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition / F. Höppner, F. Klawonn, R. Kruse, T. Runkler. – Chichester : John Wiley & Sons, 1999. – 300 p.
3. Xu R. Clustering / R. Xu, D. C. Wunsch. – Hoboken N. J.: John Wiley & Sons, Inc., 2009. – 398 p.
4. T. Kohonen. Self-Organizing Maps / T. Kohonen. – Berlin : Springer-Verlag, 1995.
5. Krishnapuram R. A Possibilistic Approach to Clustering / R. Krishnapuram, J. M. Keller // IEEE Transactions on

- Fuzzy Systems, May 1993: Proceedings. – IEEE, 1993. – Vol. 1. – P. 98–110. DOI: 10.1109/91.227387.
6. Bodyanskiy, Ye. Computational intelligence techniques for data analysis / Ye. Bodyanskiy // Lecture Notes in Informatics. – Bonn : Gesellschaft für Informatik, 2005. – P. 15–36.
7. Grosan C. Swarm intelligence in Data Mining / C. Grosan, A. Abraham, M. Chis // Studies in Computational Intelligence. – 2006. – №34. – P. 1–20.
8. Chu S.-C. Cat swarm optimization / S.-C. Chu, P.-W. Tsai, J. S. Pan // Lecture Notes in Artificial Intelligence. – 4099. – Berlin Heidelberg : Springer-Verlag, 2006. – P. 854–858.
9. Chu S.-C. Computational Intelligence based on the behavior of cats/ S.-C. Chu, P.- W. Tsai // International Journal of Innovative Computing, Information, and Control. – 2007. – Vol. 3, № 1. – P. 163–173.
10. Online neuro fuzzy clustering of data with omissions and outliers based on completion strategy [Electronic resource]/ [A. Shafronenko, Ye. Bodyanskiy, D. Rudenko] // Proceedings of The Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019), 2019. – Zaporizhzhia, 2019. – P. 18–27.
11. Shafronenko A. Yu. The Fast Modification of Evolutionary Bioinspired Cat Swarm Optimization Method [Electronic resource] / A. Yu. Shafronenko, Ye. V. Bodyanskiy, I. P. Pliss // 2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL), 2019. – Sozopol, Bulgaria, 2019. – P. 548–552. DOI: 10.1109 /CAOL46282.2019.9019583

Received 16.02.2021.

Accepted 23.04.2021.

УДК: 004.8:004.032.26

### ОНЛАЙН МЕТОД ВОЗМОЖНОСТНОЙ КЛАСТЕРИЗАЦИИ ДАННЫХ НА ОСНОВЕ ЭВОЛЮЦИОННОЙ ОПТИМИЗАЦИИ КОШАЧЬИХ СТАЙ

**Бодянский Е. В.** – д-р техн. наук, профессор, профессор кафедры искусственного интеллекта Харьковского национального университета радиоэлектроники, Харьков, Украина.

**Шафроненко А. Ю.** – канд. техн. наук, доцент, доцент кафедры информатики Харьковского национального университета радиоэлектроники, Харьков, Украина.

**Климова И. Н.** – ассистент кафедры системной инженерии Харьковского национального университета радиоэлектроники, Харьков, Украина.

### АННОТАЦИЯ

**Актуальность.** Проблемы кластеризации больших данных на сегодняшний день являются актуальной сферой искусственного интеллекта. Данная задача часто встречается во многих приложениях, связанных с интеллектуальным анализом данных, глубоким обучением и тому подобное. Для решения данных проблем традиционные подходы и методы требуют, чтобы вся выборка данных подавалась в пакетной форме.

**Цель.** Целью работы было предложить метод нечеткой вероятностной кластеризации данных с использованием эволюционной оптимизации кошачьих стай, который был бы лишен недостатков традиционных подходов кластеризации данных.

**Метод.** Процедура нечеткой вероятностной кластеризации данных с использованием эволюционных алгоритмов для быстрого определения экстремумов целевой функции, центроидов кластеров и адаптивных функций принадлежности, позволяющих не тратить ресурсы машины для хранения промежуточных расчетов и не требующих дополнительного времени для решения поставленной задачи кластеризации данных, независимо от размерности и способа подачи на обработку.

**Результаты.** Предложенный алгоритм кластеризации данных на основе эволюционной оптимизации достаточно прост в численной реализации, лишен недостатков, присущих традиционным методам нечеткой кластеризации, и может работать с большим объемом входной информации, обрабатываемой в режиме онлайн в реальном времени.

**Выводы.** Результаты эксперимента позволяют рекомендовать разработанный метод для решения проблем автоматической кластеризации и классификации больших данных, максимально быстро находить экстремумы выборки, независимо от способа подачи данных на обработку. Предложенный метод онлайн возможностной нечеткой кластеризации данных на основе эволюционной оптимизации кошачьих стай предназначен для использования в гибридных системах вычислительно-интеллекта, нейро-фаззи системах, в обучении искусственных нейронных сетей, в задачах кластеризации и классификации.

**КЛЮЧЕВЫЕ СЛОВА:** нечеткая кластеризация, правило обучения, оптимизация кошачьей стаи, режим трассировки, режим поиска.

UDC 004.8:004.032.26

## ONLINE PROBABILISTIC FUZZY CLUSTERING METHOD BASED ON EVOLUTIONARY OPTIMIZATION OF CAT SWARM

**Bodyanskiy Ye. V.** – Dr. Sc., Professor at the Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

**Shafronenko A. Yu.** – PhD, Associated Professor at the Department of Informatics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

**Klymova I. N.** – Assistant at the Department of System Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

### ABSTRACT

**Context.** The problems of big data clustering today is a very relevant area of artificial intelligence. This task is often found in many applications related to data mining, deep learning, etc. To solve these problems, traditional approaches and methods require that the entire data sample be submitted in batch form.

**Objective.** The aim of the work is to propose a method of fuzzy probabilistic data clustering using evolutionary optimization of cat swarm, that would be devoid of the drawbacks of traditional data clustering approaches.

**Method.** The procedure of fuzzy probabilistic data clustering using evolutionary algorithms, for faster determination of sample extrema, cluster centroids and adaptive functions, allowing not to spend machine resources for storing intermediate calculations and do not require additional time to solve the problem of data clustering, regardless of the dimension and the method of presentation for processing.

**Results.** The proposed data clustering algorithm based on evolutionary optimization is simple in numerical implementation, is devoid of the drawbacks inherent in traditional fuzzy clustering methods and can work with a large size of input information processed online in real time.

**Conclusions.** The results of the experiment allow to recommend the developed method for solving the problems of automatic clustering and classification of big data, as quickly as possible to find the extrema of the sample, regardless of the method of submitting the data for processing. The proposed method of online probabilistic fuzzy data clustering based on evolutionary optimization of cat swarm is intended for use in hybrid computational intelligence systems, neuro-fuzzy systems, in training artificial neural networks, in clustering and classification problems.

**KEYWORDS:** fuzzy clustering, learning rule, cat swarm optimization, tracing mode, seeking mode.

### REFERENCES

1. Bezdek J. C. Pattern recognition with fuzzy objective function algorithms. New York, Springer, 1981, 253 p. DOI <https://doi.org/10.1007/978-1-4757-0450-1>.
2. Höppner F., Klawonn F., Kruse R., Runkler T. Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition. Chichester, John Wiley & Sons, 1999, 300 p.
3. Xu R., Wunsch D. C. Clustering. Hoboken N. J., John Wiley & Sons, Inc., 2009, 398 p.
4. Kohonen T. Self-Organizing Maps. Berlin, Springer-Verlag, 1995.
5. Krishnapuram R., Keller J. M. A Possibilistic Approach to Clustering, *IEEE Transactions on Fuzzy Systems*, May 1993: *Proceedings*, IEEE, 1993, Vol. 1, pp. 98–110. DOI: 10.1109/91.227387.
6. Bodyanskiy, Ye. Computational intelligence techniques for data analysis, *Lecture Notes in Informatics*. Bonn, Gesellschaft für Informatik, 2005, pp. 15–36.
7. Grosan C., Abraham A., Chis M. Swarm intelligence in Data Mining, *Studies in Computational Intelligence*, 2006, No. 34, pp. 1–20.
8. Chu S.-C., Tsai P.-W., Pan J. S. Cat swarm optimization, *Lecture Notes in Artificial Intelligence*, 4099. Berlin Heidelberg, Springer-Verlag, 2006, pp. 854–858.
9. Chu S.-C., Tsai P.-W. Computational Intelligence based on the behavior of cats, *International Journal of Innovative Computing, Information, and Control*, 2007, Vol. 3, No. 1, pp. 163–173.
10. Shafronenko A., Bodyanskiy Ye., Rudenko D. Online neuro fuzzy clustering of data with omissions and outliers based on completion strategy [Electronic resource], *Proceedings of The Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019)*, 2019. Zaporizhzhia, 2019, pp. 18–27.
11. Shafronenko A. Yu., Bodyanskiy Ye. V., Pliss I. P. The Fast Modification of Evolutionary Bioinspired Cat Swarm Optimization Method [Electronic resource], *2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL)*, 2019. Sozopol, Bulgaria, 2019, pp. 548–552. DOI: 10.1109/CAOL46282.2019.9019583