

ПРАВДОПОДІБНА НЕЧІТКА КЛАСТЕРИЗАЦІЯ ДАНИХ НА ОСНОВІ ЕВОЛЮЦІЙНОГО МЕТОДУ БОЖЕВІЛЬНИХ КОТІВ

Є.В. БОДЯНСЬКИЙ, А.Ю. ШАФРОНЕНКО, І.П. ПЛИСС

Анотація. Розглянуто задачу нечіткої кластеризації великих масивів, що подаються на опрацювання як у пакетному, так і в онлайн режимах на основі правдоподібного підходу. Для відшукування глобального екстремуму цільової функції правдоподібної нечіткої кластеризації введено модифікацію роївого алгоритму зграй божевільних котів, яка об'єднує в собі переваги еволюційних алгоритмів та глобального випадкового пошуку. Показано, що різні режими пошуку породжуються уніфікованою математичною процедурою, окремими випадками якої є відомі алгоритми як локальної, так і глобальної оптимізації. Запропонований підхід є досить простим в обчислювальній реалізації і характеризується високою швидкістю та надійністю у задачах багатоекстремальної нечіткої кластеризації.

Ключові слова: нечітка кластеризація, теорія правдоподібності, еволюційні методи оптимізації, правдоподібна нечітка кластеризація, центроїди-прототипи, котяча зграя, режим трасування, режим пошуку, рівень належності.

ВСТУП

Задача кластеризації є важливою частиною Data Mining, основною відмінністю якої є те, що в її основу покладено парадигму самонавчання, тобто відсутність заздалегідь розміченої навчальної вибірки [1,2]. У межах цієї задачі сформувалася множина підходів для її розв'язання, проте з практичного погляду найбільш ефективними є методи нечіткої кластеризації, в основі яких лежить припущення, що формуються класи-кластери, які взаємно перетинаються, тобто одне і те ж спостереження (вектор або матриця) можуть одночасно належати до декількох класів.

У межах нечіткої кластеризації склалися два основні напрями: імовірнісні та можливісні методи [3, 4], які підтвердили свою ефективність розв'язання множини реальних задач, проте не позбавлені і низки недоліків. Це передусім чутливість імовірнісних алгоритмів до наявності викривлених збуреннями та викидами спостережень (алгоритми не мають робастних властивостей), проблема збіжності в можливісній кластеризації, обчислювальні проблеми під час роботи з даними високої розмірності [4, 5], багатоекстремальність цільових функцій [6–9], яка спричиняє зупинку процесу кластеризації в локальних екстремумах. Деякі з цих проблем можуть бути вирішені за допомогою правдоподібної нечіткої кластеризації [10, 11], основу якої

становить теорія довіри [12], однак проблема багатоекстремальності все одно залишається. Для подолання цієї проблеми використання алгоритмів генетичного пошуку та методу нечітких J-середніх [9], що є за суттю локальним евристичним пошуком, у загальному випадку не гарантує відшукування глобального екстремуму. У цій ситуації перспективним видається використання так званого ройового інтелекту [13] і алгоритмів оптимізації на основі роїв частинок [14], серед яких вельми ефективними показали себе оптимізаційні методи котячих зграй [15–19] та їх різні модифікації [20, 22]. Пришвидшити процес пошуку можна, використовуючи більш просунуті версії котячих методів оптимізації, такі як оптимізаційні методи божевільних котів (crazy cats) [23, 24] у поєднанні з математичним апаратом випадкового глобального пошуку [25, 26]. Цікавим є також поєднання ідей теорії правдоподібності з методами глобальної оптимізації на основі еволюційного алгоритму божевільних котів для розв’язання широкого класу задач нечіткої кластеризації.

ЗМІСТОВНА ПОСТАНОВКА ЗАВДАННЯ

Вихідною інформацією для розв’язання задачі кластеризації є вибірка, що містить N n -вимірних векторів спостережень $X = \{x_1, \dots, x_k, \dots, x_N\}$, яка повинна бути розбита на m взаємно перетинних класів-кластерів. У процесі кластеризації мають бути знайдені центроїди-прототиби цих кластерів w_q , $q = 1, 2, \dots, m$; рівні належності кожного спостереження x_k до кожного кластера $Cl_q - U_q(k)$ і оцінювання правдоподібності $Cr_q(k)$ того, що x_k дійсно належить до Cl_q . При цьому важливо відзначити, що у загальному випадку $U_q(k) \neq Cr_q(k)$.

ПРАВДОПОДІБНА НЕЧІТКА КЛАСТЕРИЗАЦІЯ ІЗ ФУНКЦІЯМИ НАЛЕЖНОСТІ СПЕЦІАЛЬНОГО ТИПУ

Математично завдання правдоподібної нечіткої кластеризації пов’язане з мінімізацією цільової функції

$$E(Cr_q(k), w_q) = \sum_{k=1}^N \sum_{q=1}^m Cr_q^\beta(k) D^2(x(k), w_q) \quad (1)$$

за обмежень

$$\begin{cases} 0 \leq Cr_q(k) \leq 1 \quad \forall q, k, \\ \sup Cr_q(k) \geq 0,5 \quad \forall k, \\ Cr_q(k) + \sup Cr_l(k) = 1, \quad l = 1, 2, \dots, m; \quad l \neq q \end{cases}$$

для всіх q і k , у яких $Cr_q(k) \geq 0,5$. Тут $\beta > 1$ — фазифікатор,

$D^2(x_k, w_q) = \|x_k - w_q\|^2$ — звичайна евклідова відстань.

Треба зазначити, що цільова функція (1) майже збігається із цільовою функцією імовірнісної нечіткої кластеризації [4] з тією різницею, що в ній замість рівнів належності $U_q(k)$ використовуються рівні довіри $Cr_q(k)$, що дозволяє відмовитись від імовірнісних обмежень вигляду

$$\sum_{q=1}^m U_q(k) = 1.$$

Зазначимо також, що за умови $\beta = 2$ (1) є близькою до цільової функції популярного методу нечітких c -середніх (FCM) Дж. Бездека [3].

Для оцінювання рівнів належності в правдоподібному підході використовується функція належності вигляду

$$U_q(k) = \phi_q(D(x_k, w_q)),$$

яка є за суттю мірою подібності, що використовує оцінку відстані $D(x_k, w_q)$.

У працях [10, 11] як таку функцію запропоновано використовувати вираз

$$U_q(k) = \frac{1}{1 + D^2(x_k, w_q)}, \quad (2)$$

який є дзвонуватою функцією щільності розподілу Коші з одиничним параметром ширини. Можна показати, що використання оцінок імовірнісної нечіткої належності

$$\begin{aligned} U_q(k) &= \frac{(D^2(x_k, w_q))^{-\frac{1}{\beta}}}{\sum_{l=1}^m (D^2(x_k, w_l))^{-\frac{1}{\beta}}} = \frac{(D^2(x_k, w_q))^{-\frac{1}{\beta}}}{(D^2(x_k, w_q))^{-\frac{1}{\beta}} + \sum_{\substack{l=1 \\ l \neq q}}^m (D^2(x_k, w_l))^{-\frac{1}{\beta}}} = \\ &= \frac{1}{1 + \frac{\sum_{\substack{l=1 \\ l \neq q}}^m (D^2(x_k, w_l))^{-\frac{1}{\beta}}}{(D^2(x_k, w_q))^{-\frac{1}{\beta}}}} \end{aligned}$$

для $\beta = 2$ дозволяє отримати розподіл Коші

$$U_q(k) = \frac{1}{1 + \frac{\|x_k - w_q\|^2}{\sigma_q^2}}$$

з параметром ширини

$$\sigma_q^2 = \frac{1}{\sum_{\substack{l=1 \\ l \neq q}}^m \|x_k - w_l\|^{-2}}, \quad (3)$$

що є узагальненням (2) і збігається з ним за умови $\sigma_q^2 = 1$. Таким чином, рівні належності в правдоподібному підході збігаються з імовірнісними рівнями належності, які використовуються у FCM.

Таким чином, модифікований варіант процедури правдоподібної кластеризації може бути записаний у вигляді:

$$\left\{ \begin{array}{l} U_q(k) = \frac{1}{1 + \frac{\|x_k - w_q\|^2}{\sigma_q^2}}, \\ U_q^*(k) = \frac{U_q(k)}{\sup U_l(k)}, \\ Cr_q(k) = \frac{1}{2} \left(U_q(k) + 1 - \sup_{l \neq q} U_l^*(k) \right), \\ w_q = \frac{\sum_{k=1}^N (Cr_q(k))^\beta x_k}{\sum_{k=1}^N (Cr_q(k))^\beta}, \end{array} \right. \quad (4)$$

де σ_q^2 визначається виразом (3).

Якщо ж дані надходять на оброблення послідовно в онлайн режимі, можна скористатися адаптивним правдоподібним методом кластеризації у вигляді:

$$\left\{ \begin{array}{l} \sigma_q^2(k+1) = \frac{1}{\sum_{\substack{l=1 \\ l \neq q}}^m \|x_{k+1} - w_l(k)\|^{-2}}, \\ U_q(k+1) = \frac{1}{1 + \frac{\|x_{k+1} - w_q(k)\|^2}{\sigma_q^2(k+1)}}, \\ U_q^*(k+1) = \frac{U_q(k+1)}{\sup U_l(k+1)}, \\ Cr_q(k+1) = \frac{1}{2} \left(U_q^*(k+1) + 1 - \sup_{l \neq q} U_l^*(k+1) \right), \\ w_q(k+1) = w_q(k) + \eta(k+1) Cr_q^\beta(k+1) (x_{k+1} - w_q(k)), \end{array} \right. \quad (5)$$

де $\eta(k+1)$ — параметр кроку навчання.

Тут необхідно відзначити, що процедура (4) є результатом пакетної оптимізації цільової функції (1), а (5) — градієнтним алгоритмом пошуку екстремуму цієї ж функції, тобто гарантується відшукання тільки локального екстремуму. Водночас у працях [6, 7] показано, що задачу умовної оптимізації цільової функції ймовірнісної нечіткої кластеризації можна подати у вигляді задачі безумовної оптимізації функції

$$E(w_q) = \sum_{k=1}^N \left(\sum_{q=1}^m \|x_k - w_q\|^{2(1-\beta)} \right)^{1-\beta}$$

або за умови $\beta = 2$:

$$E(w_q) = \sum_{k=1}^N \left(\sum_{q=1}^m \|x_k - w_q\|^{-2} \right)^{-1},$$

що має множину локальних екстремумів. Тому можна говорити про відшукування лише одного з них, найближчого до початкової точки пошуку. Саме таким екстремумом є ройові алгоритми оптимізації, які повинні бути відповідним чином адаптовані до розглянутої задачі і відповідати вимогам високої швидкодії.

АЛГОРИТМ ГЛОБАЛЬНОЇ ОПТИМІЗАЦІЇ БОЖЕВІЛЬНОЇ КОТЯЧОЇ ЗГРАЇ В ЗАДАЧІ НЕЧІТКОЇ КЛАСТЕРИЗАЦІЇ

Для пошуку глобального екстремуму цільової функції нечіткої кластеризації пропонується використовувати модифікований метод оптимізації божевільної котячої зграї, синтезований на основі оптимізаційного підходу котячої зграї [16–24] і методів глобального випадкового пошуку [25, 26].

Ідея оптимізації на основі еволюційного алгоритму котячої зграї полягає в тому, що формується група-зграя котів, кожен з яких рухається у напрямку або локального, або глобального екстремуму прийнятої цільової функції $E(w_q)$. При цьому ця зграя складається з Q особин cat_p , $p = 1, 2, \dots, Q$, кожна з яких може перебувати в одному з двох можливих станів: режим пошуку (SM) локальних екстремумів і режим погоні (TM), що ставить собі за мету відшукування глобального екстремуму; SM, як правило, реалізується на основі градієнтного пошуку з малим параметром навчання і таким чином сканує локальний окіл кожного з котів, що міститься в цьому режимі, TM характеризується випадковими стрибками з великою амплітудою і ставить собі за мету «витягти» kota cat_p з локального екстремуму в разі його потрапляння туди.

У загальному випадку стандартному алгоритму котячої зграї можна надати вигляду послідовності ітерацій [15, 16]:

CS1: випадковим чином створюється зграя з Q котів cat_p , кожен з яких є за суттю n -вимірним вектором $w_p(0)$ в області визначення функції, що оптимізується, та оцінюється значення функції в цій точці $E(w_p(0))$.

CS2: вводиться параметр стану SPC, що набуває значення 0 або 1, за допомогою якого вихідна зграя розбивається випадковим чином на дві групи: якщо $SPC = 1$, то кіт перебуває в режимі пошуку, якщо ж $SPC = 0$, то відповідний кіт перебуває в режимі погоні.

CS3: коти з $SPC = 1$ починають пошук локального екстремуму, а коти з $SPC = 0$ запускаються в режим погоні.

CS4: оцінюються значення цільової функції для всіх котів і зберігаються всі $w_p(1)$ з найменшими значеннями цієї функції, коти з найбільшим значенням $E(w_p(1))$ можуть бути видалені зі зграї.

CS5: повернення до CS1 з новими значеннями, тобто починається новий етап з оновленою популяцією.

У загальному випадку обидва режими SM і TM реалізуються паралельно, при цьому SM фактично базується на основі покоординатного спуску, тобто в кожен конкретний момент часу може змінюватися тільки одна координата n -вимірного простору пошуку, що природно знижує швидкодію процедури. У режимі погоні швидкості руху по кожній координаті також оцінюються незалежно одна від одної, що знову-таки знижує швидкодію.

Для подолання цих недоліків у праці [21] був запропонований рандомізований алгоритм оптимізації на основі котячих зграй, що забезпечує підвищену швидкодію порівняно з відомою процедурою-прототипом. При цьому рух kota в режимі пошуку можна описати за допомогою рекурентної процедури

$$w_p(\tau + 1) = w_p(\tau) - \eta_{SM} \hat{\nabla} E(w_p(\tau)),$$

де $\hat{\nabla} E(w_p(\tau))$ — оцінка градієнта функції, що оптимізується в точці $w_p(\tau)$, одержувана або на основі пошуку з центральною пробою [25], або на основі випадкових проб (статистичний градієнт [26]); η_{SM} — малий крок пошуку в просторі R^n .

Рух kota в режимі погоні описується алгоритмом, що є гібридом популярного методу оптимізації важкої кульки і випадкового пошуку [22]:

$$w_p(\tau + 1) = w_p(\tau) - \alpha(w_p(\tau) - w_p(\tau - 1)) - \eta_{TM} \hat{\nabla} E(w_p(\tau)) + \Xi(\tau), \quad (6)$$

де $0 < \alpha < 1$ — параметр інерції режиму погоні; $\Xi(\tau)$ — випадкове збурення, що вводить додаткове сканування простору пошуку.

У працях [23, 27] для поліпшення процесу пошуку глобального екстремуму в режимі погоні в алгоритм руху кожного kota додатково був уведений «фактор божевільності», який описується набором випадкових параметрів і дозволяє здійснювати раптові стрибки, що змінюють траєкторію руху шляхом варіювання характеристик сигналу збурення $\Xi(\tau)$.

Для керування сигналом $\Xi(\tau)$ доцільно скористатися ідеєю блукаючого глобального пошуку [26], який довів свою ефективність розв'язання багатоекстремальних задач.

При цьому характеристики випадкового збурення Л. Растрігін [25, 26] запропонував змінювати відповідно до виразу

$$\Xi(\tau) = \gamma \Xi(\tau - 1) - \delta(E(w_p(\tau)) - E(w_p(\tau - 1))) + \sigma^2 H(k), \quad (7)$$

де γ — параметр коригування характеристик збурення; $0 \leq \delta \leq 1$ — параметр швидкості самонавчання типу параметра інерції α у праці (6); σ^2 — дисперсія білого шуму $H(\tau)$.

Таким чином, весь процес оптимізації за допомогою підходу божевільних котів можна описати за допомогою рекурентних співвідношень (6), (7), при цьому, якщо $\alpha = \gamma = \delta = 0$, режим погоні автоматично переходить у режим пошуку (рух по антиградієнту).

РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ

Щоб перевірити ефективність та оцінити працездатність і спроможність якісно кластеризувати великі обсяги даних розробленого методу, а також довести його перевагу над аналогами, проведено експериментальне дослідження за допомогою чотирьох різних навчальних вибірок, а саме: Іриси Фішера, Cancer, Вина та Glass [5] (табл. 1), параметричні характеристики яких наведено в табл. 2.

Таблиця 1. Характеристики навчальних вибірок

Навчальна вибірка	Кількість кластерів	Кількість атрибутів	Кількість спостережень
Іриси Фішера	3	4	150
Cancer	2	9	683
Вина	3	13	178
Glass	6	8	214

Таблиця 2. Порівняльні результати показників ефективності таких алгоритмів, як PSO, CSO

Навчальна вибірка	MSE	PSO	CSO	Правдоподібна нечітка кластеризація даних на основі еволюційного методу божевільних котів
Іриси Фішера	Найкраще	$4 \cdot 10^{-7}$	$9 \cdot 10^{-10}$	$8,4 \cdot 10^{-11}$
	Середнє	$7 \cdot 10^{-6}$	$7,3 \cdot 10^{-9}$	$5,2 \cdot 10^{-10}$
	Найгірше	$1,2 \cdot 10^{-5}$	$9,3 \cdot 10^{-9}$	$9,6 \cdot 10^{-10}$
Cancer	Найкраще	$1,3 \cdot 10^{-7}$	$8 \cdot 10^{-10}$	$8,5 \cdot 10^{-11}$
	Середнє	$7 \cdot 10^{-6}$	$4,4 \cdot 10^{-9}$	$7,6 \cdot 10^{-11}$
	Найгірше	$2,02 \cdot 10^{-5}$	$6,8 \cdot 10^{-9}$	$7,8 \cdot 10^{-10}$
Вина	Найкраще	$1,4 \cdot 10^{-6}$	$7 \cdot 10^{-10}$	$9,5 \cdot 10^{-11}$
	Середнє	$5 \cdot 10^{-5}$	$4 \cdot 10^{-9}$	$6,6 \cdot 10^{-11}$
	Найгірше	$2 \cdot 10^{-5}$	$6 \cdot 10^{-9}$	$6,8 \cdot 10^{-10}$
Glass	Найкраще	$2,5 \cdot 10^{-7}$	$7,9 \cdot 10^{-10}$	$8,7 \cdot 10^{-11}$
	Середнє	$6,9 \cdot 10^{-6}$	$5 \cdot 10^{-9}$	$7,7 \cdot 10^{-11}$
	Найгірше	$3 \cdot 10^{-5}$	$5,9 \cdot 10^{-9}$	$7,7 \cdot 10^{-10}$

Аналізуючи та оцінюючи отримані результати експериментальних досліджень, можна зробити висновки, що запропонований метод правдоподібної нечіткої кластеризації даних на основі еволюційного методу божевіль-

них котів забезпечує достатньо якісні результати кластеризації, що підтверджується експериментально.

ВИСНОВКИ

Запропоновано підхід до розв'язання багатоекстремальної задачі правдоподібної нечіткої кластеризації на основі модифікованої оптимізаційної процедури божевільних котів. Особливістю запропонованої модифікованої процедури оптимізації є можливість керування випадковими збуреннями, що визначають властивості режиму трасування (стеження) котячої зграї. Відзначено, що і режим пошуку, і режим погоні формально можна описати в межах єдиної процедури з різними параметрами алгоритму оптимізації, синтезованого на основі випадкової зграї божевільних котів у поєднанні з випадковим глобальним пошуком.

Запропонований підхід є простим в обчислювальній реалізації і забезпечує високу швидкодію розв'язання задач нечіткої кластеризації.

ЛІТЕРАТУРА

1. R. Xu and D.C. Wunsch, *Clustering*. Hoboken, N.J.: John Wiley & Sons, Inc., 2009.
2. C.C. Aggarwal, *Data Mining: Text Book*. Springer, 2015.
3. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. N.Y.: Plenum Press, 1981.
4. F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition*. Chichester: John Wiley & Sons, 1999.
5. *UCI Machine Learning Repository, Data Sets*. Available at: <https://archive.ics.uci.edu/>
6. R.J. Hathaway and J.C. Bezdek, "Optimization of clustering criteria by reformulation", *IEEE Transactions on Fuzzy Systems*, vol. 3, pp. 241–245, 1995.
7. N.R. Pal, J.C. Bezdek, and R.J. Hathaway, "Sequential Competitive Learning and the Fuzzy C-Means Clustering Algorithms", *Neural Networks*, vol. 9, no. 5, pp. 787–796, 1996.
8. P. Hansen and N. Mladenović, "J-Means: A new local search heuristic for minimum sum-of-squares clustering", *Pattern Recognition*, vol. 34, pp. 405–413, 2001.
9. N. Belacel, P. Hansen, and N. Mladenović, "Fuzzy J-Means: A new heuristic for fuzzy clustering", *J. Pattern Recognition*, vol. 35, pp. 2193–2200, 2002.
10. J. Zhou, Q. Wang, C.-C. Hung, and X. Yi, "Credibilistic clustering: the model and algorithms", *Int.J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 23, no. 4, pp. 545–564, 2015.
11. J. Zhou, Q. Wang, C.-C. Hung, "Credibilistic clustering algorithms via alternating cluster estimation", *J. Intell. Manuf.*, vol. 28, pp. 727–738, 2017.
12. B. Liu and Y. Liu, "Expected value of fuzzy variable and fuzzy expected value models", *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 4, pp. 445–450, 2002.
13. C. Grosan, A. Abraham, and M. Chis, "Swarm intelligence in data mining", *Swarm Intelligence and Data Mining*. Springer, Germany, 2006.
14. J. Kennedy and R. Eberhart, "Particle swarm optimization", *Proc. IEEE Int. Conf. on Neural Networks*, vol. 4, pp. 1942–1948, 1995.
15. S.C. Chu, P.W. Tsai, and J.S. Pan, "Cat swarm optimization", *Lecture Notes in Artificial Intelligence*, vol. 4099, pp. 854–858. Berlin Heidelberg, Springer-Verlag, 2006.

16. S.C. Chu and P.W. Tsai, “Computational Intelligence based on the behavior of cats”, *Int. J. of Innovative Computing, Information, and Control*, vol. 3, no. 1, pp. 163–173, 2007.
17. G. Panda, P.M. Pradhan, and B. Majhi, “Direct and inverse modeling of plants using cat swarm optimization”, *Handbook of Swarm Intelligence, ALO 8*; eds. B.K. Panigrahi, Y. Shi, M-H. Zim. Springer-Verlag, Berlin Heidelberg, 2011, pp. 469–485.
18. M. Orouskhani, Y. Orouskhani, and M. Teshnehlab, “Average-inertia weighted cat swarm optimization”, *Lecture Notes in Computing Science*. Springer-Verlag, Berlin Heidelberg, 2011, pp. 321–328.
19. M. Orouskhani, Y. Orouskhani, M. Mansouri, and M. Teshnehlab, “A novel cat swarm optimization algorithm for unconstrained optimization problems”, *Int. J. Information Technology and Computer Science*, vol. 11, pp. 32–41, 2013.
20. Y. Zhang and Y. Tian, “An improved cat swarm optimization algorithm and application research”, *Proc. 7th Int. Conf. on Advanced Computational Intelligence*, Mount Wuyi, Fujian, China 2015, pp. 207–211.
21. A.Yu. Shafronenko, Ye.V. Bodyanskiy, and I.P. Pliss, “The Fast Modification of Evolutionary Bioinspired Cat Swarm Optimization Method”, *Proc. 2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL), 2019, Sozopol, Bulgaria*, pp. 548–552. doi: 10.1109 /CAOL46282. 2019.9019583.
22. A. Shafronenko and Ye. Bodyanskiy, “Adaptive fuzzy clustering approach based on evolutionary cat swarm optimization”, *Proc. Third International Workshop on Computer Modeling and Intelligent Systems — CMIS, 2020, Zaporizhzhia, Ukraine*, pp. 832–842.
23. S.K. Saha, R. Kar, D. Mandal, and S.P. Ghoshal, “IIR filter design with craziness based particle swarm optimization technique”, *World Academy of Science, Engineering and Technology*, pp.1628–1635, 2011.
24. A. Sarangi, S.K. Sarangi, M. Mukherjee, and S.P. Panigrahi, “System identification by crazy-cat swarm optimization”, *Proc. Int. Conf. on Microwave, Optical and Communication Engineering, Bhubaneswar, India- 2015*, pp. 439–442.
25. L.A. Rastrigin, *Statistical search methods* [in rus.]. Moscow: Science, 1968.
26. L.A. Rastrigin, *Extreme Control Systems* [in rus.]. Moscow: Science, 1974.
27. Jian Zhou and Chih-Cheng Hung, *A Generalized Approach to Possibilistic Clustering Algorithms*. Faculty Publications, 2007.
28. F.W. Young and R.M. Hamer, *Theory and Applications of Multidimensional Scaling-Hillsdale*. N.J.: Erlbaum, 1994.
29. A. Shafronenko, Ye. Bodyanskiy, I. Klymova, and O. Holovin, “Online credibilistic fuzzy clustering of data using membership functions of special type”, *Proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020)*, April 27–1 May 2020, Zaporizhzhia, pp. 744–753.

Надійшла 01.07.2021

INFORMATION ON THE ARTICLE

Yevgeniy V. Bodyanskiy, ORCID: 0000-0001-5418-2143, Kharkiv National University of Radio Electronics, Ukraine, e-mail: yevgeniy.bodyanskiy@nure.ua

Alina Yu. Shafronenko, ORCID: 0000-0002-8040-0279, Kharkiv National University of Radio Electronics, Ukraine, e-mail: alina.shafronenko@nure.ua

Iryna P. Pliss, ORCID: 0000-0001-7918-7362, Kharkiv National University of Radio Electronics, Ukraine, e-mail: iryna.pliss@nure.ua

ПРАВДОПОБНАЯ НЕЧЕТКАЯ КЛАСТЕРИЗАЦИЯ ДАННЫХ НА ОСНОВЕ ЭВОЛЮЦИОННОГО МЕТОДА БЕЗУМНЫХ КОТОВ / Е.В. Бодянский, А.Ю. Шафроненко, И.П. Плисс

Аннотация. Рассмотрена задача нечеткой кластеризации больших массивов, которые подаются на обработку как в пакетном, так и онлайн режимах на основе правдоподобного подхода. Для отыскания глобального экстремума целевой функции правдоподобной нечеткой кластеризации введена модификация роевого алгоритма стай безумных котов, которая объединяет в себе преимущества эволюционных алгоритмов и глобального случайного поиска. Показано, что различные режимы поиска порождаются унифицированной математической процедурой, частными случаями которой являются известные алгоритмы как локальной, так и глобальной оптимизации. Предложенный подход является достаточно простым в вычислительной реализации и характеризуется высоким быстродействием и надежностью в задачах многоэкстремальной нечеткой кластеризации.

Ключевые слова: нечеткая кластеризация, теория правдоподобности, эволюционные методы оптимизации, правдоподобная нечеткая кластеризация, центры-прототипы, кошачья стая, режим трассировки, режим поиска, уровень принадлежности.

CREDIBILISTIC FUZZY CLUSTERING BASED ON EVOLUTIONARY METHOD OF CRAZY CATS / Ye.V. Bodyanskiy, A.Yu. Shafronenko, I.P. Pliss

Abstract. The problem of fuzzy clustering of large datasets that are sent for processing in both batch and online modes, based on a credibilistic approach, is considered. To find the global extremum of the credibilistic fuzzy clustering goal function, the modification of the swarm algorithm of crazy cats swarms was introduced, that combined the advantages of evolutionary algorithms and a global random search. It is shown that different search modes are generated by a unified mathematical procedure, some cases of which are known algorithms for both local and global optimizations. The proposed approach is easy to implement and is characterized by the high speed and reliability in problems of multi-extreme fuzzy clustering.

Keywords: fuzzy clustering, credibility theory, evolutionary methods of optimization, credibilistic fuzzy clustering, centroids-prototypes, cats swarm, tracing mode, seeking mode, membership level.