

УДК 004.032.26

[0000-0003-3841-7373] **Є. Є. Федоров**, д-р техн. наук, професор,
[0000-0002-6614-4133] **Т. Ю. Уткіна**, канд. техн. наук, доцент
e-mail: t.utkina@chdtu.edu.ua

Черкаський державний технологічний університет
б-р Шевченка, 460, м. Черкаси, 18006, Україна

МЕТОД КЛАСТЕРИЗАЦІЇ КВАЗІПЕРІОДИЧНОГО СИГНАЛУ НА ОСНОВІ АЛГОРИТМУ КЛОНАЛЬНОГО ВІДБОРУ

У роботі запропоновано ієрархічно-ітеративний метод кластеризації квазіперіодичного сигналу, що базується на алгоритмі клонального відбору, підвищує швидкість та точність кластеризації. Попередньо проводиться перетворення зразків (квазіперіодичних ділянок) цього сигналу до єдиного амплітудно-часового вікна на основі зсуву та масштабування за часом і амплітудою, лінійною інтерполяцією та дискретизацією за часом.

У результаті проведеного аналізу сучасних методів кластеризації квазіперіодичних сигналів встановлено, що більшість із них має один або кілька з таких недоліків: невідома точна кількість кластерів; чутливість до початкових значень центроїдів кластерів; низька ймовірність кластеризації; низька швидкість кластеризації; порівняння ділянок сигналу, що мають лише однаковий розмір; порівняння лише бінарних сигналів. Тому актуальною є розробка методу перетворення квазіперіодичного сигналу й ієрархічно-ітеративного методу кластеризації на основі алгоритму клонального відбору. Це забезпечить підвищення ефективності аналізу структури квазіперіодичного сигналу при проведенні обробки цифрових даних в інтелектуальних комп'ютерних системах ідентифікації особистості, технічної і медичної діагностики, аналізу мережевого трафіку тощо.

Наведено порівняння запропонованого авторами методу й існуючих методів кластеризації, при цьому параметр клонування $\alpha=0.1$, параметр мутації $\beta=2.5$, кількість замінних антитіл $d=0.2|N|$. Кластеризація проводилася на квазіперіодичних звуках мовлення, виголошених різними дикторами. Запропоновано метод перетворення квазіперіодичного сигналу, який перетворює зразки (квазіперіодичні ділянки) цього сигналу до єдиного амплітудно-часового вікна, за допомогою зсуву й масштабування за часом і амплітудою, інтерполяцією та дискретизацією. Це дозволяє порівнювати зразки сигналу різної довжини та з різним розмахом амплітуд. Створено метод кластеризації квазіперіодичного сигналу на основі ієрархічно-ітеративного підходу й алгоритму клонального відбору, що зменшує чутливість до початкових значень центроїдів кластерів за рахунок випадкового пошуку та забезпечує адаптивне налаштування кількості кластерів за рахунок ієрархічного підходу, а також підвищує ймовірність кластеризації до 0.98.

Ключові слова: квазіперіодичний сигнал, алгоритм клонального відбору, ієрархічно-ітеративна кластеризація, перетворення сигналу, аналіз структури сигналу.

Вступ. Сьогодні актуальним завданням є розробка інтелектуальних комп'ютерних систем, призначених для ідентифікації, аналізу та синтезу, зберігання цифрових сигналів (наприклад акустичних, графічних, вібраційних, електрограмних, комунікаційних та ін.) [1-2].

Водночас сигнали, що обробляються комп'ютерними системами ідентифікації особистості, технічної та медичної діагностики, аналізу мережного трафіку, часто є квазіперіодичними. У зв'язку з цим постає проблема побудови ефективних методів аналізу структури квазіперіодичного сигналу. Одним із засобів аналізу структури квазіперіодичного

сигналу є кластеризація, що є різновидом машинного навчання без вчителя.

Як правило, під завданням кластеризації розуміють необхідність розбиття на підмножини (кластери) заданого набору векторів визначеної розмірності відповідно до наперед визначеного критерію мінімізації спотворення (критерій оптимізації). При цьому елементи одного кластера мають схожі ознаки, а елементи різних кластерів значно відрізняються один від одного. Залежно від шляхів оцінювання спотворення й розрізняють методи кластеризації, які є важливим елементом для вирішення загальної проблеми обробки цифрових даних.

Аналіз останніх джерел і публікацій.

Сучасні методи кластеризації являють собою множину підходів та алгоритмів – від інтуїтивних та евристичних до суто математичних і поділяються на [3-5]:

- центроїдні (к-середніх, нечітких с-середніх, зсуву середнього значення та ін.) або медоїдні (к-медоїдів, субтрактивна кластеризація та ін.) [6-8]. Методи к-медоїдів та субтрактивну кластеризацію варто застосовувати, коли наявний шум або випадкові викиди, а методи к-середніх і нечітких с-середніх – у випадку швидкої обробки великого набору об'єктів. Методи субтрактивної кластеризації та зсуву середнього значення варто застосовувати, коли невідома кількість кластерів;

- моделі суміші розподілів (ЕМ-алгоритм та ін.) [9]. Ці методи варто застосовувати, коли кластери мають різний розмір, а набір об'єктів усіх кластерів можна описати сумішшю щільностей розподілів (суміш розподілів Гаусса та ін.);

- щільнісні (DBSCAN, OPTICS та ін.) [8]. Ці методи використовують у випадку наявного шуму та випадкових викидів, коли кластери мають різні форму і розмір, невідома кількість кластерів. Для формування кластерів використовують околицю Мура чи Неймана. Кластер є топологічно зв'язною областю;

- дивізивні (DIANA) [8] або агломеративні (центроїдного зв'язку, Варда, одиночного зв'язку, повного зв'язку, групового середнього) [8]. Ці методи варто застосовувати, коли кількість об'єктів, що кластеризуються, невелика. Формують дендрограму (дерево кластерів) на основі міжкластерної відстані;

- коннекціоністські, для формування кластерів використовують штучну нейронну мережу (SOM [10-11], сімейство ART [12] та ін.);

- метаввристичні [13-15], для формування кластерів використовують метавристику (генетичний алгоритм та ін.).

Існуючі методи кластеризації мають один або декілька з таких недоліків:

- невідома точна кількість кластерів;
- чутливість до початкових значень центроїдів кластерів;
- низька ймовірність кластеризації;
- низька швидкість кластеризації;
- порівняння ділянок сигналу, що мають лише однаковий розмір;
- порівняння лише бінарних сигналів.

З метою усунення більшості з них для аналізу структури квазіперіодичного сигналу пропонується виконати його кластеризацію шляхом застосування алгоритму клонального відбору [16-18]. Цей алгоритм (рисунк 1) належить до особливого класу штучних імунних систем та моделює роботу антитіл імунної системи щодо протидії проникаючому антигену завдяки вивченню його особливостей і адаптації до них.

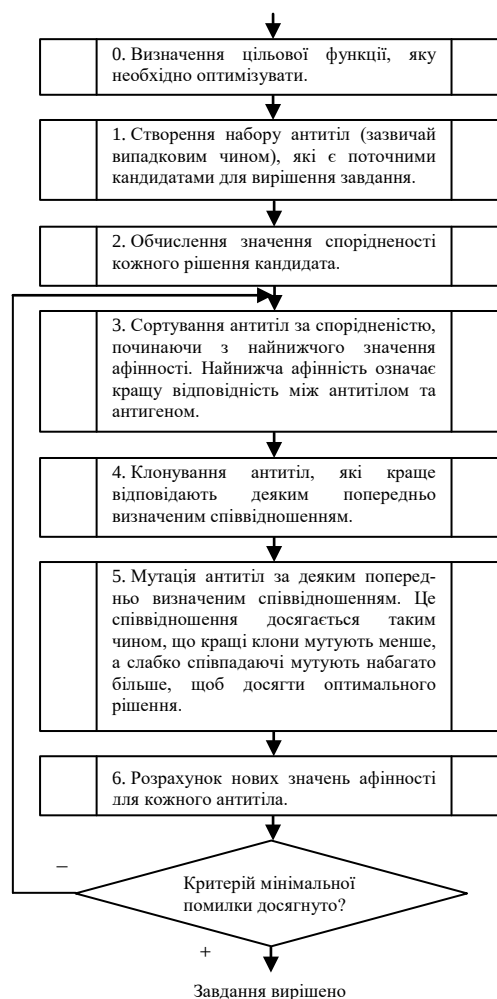


Рисунок 1 – Основні етапи алгоритму клонального відбору

Він базується на алгоритмі CLONALG, вперше запропонованому Де Кастро і Ван Зубеном [19].

На відміну від генетичного алгоритму [20], алгоритм клонального відбору використовує заміну антитіл випадково згенерованими антитілами, що підвищує різноманітність популяції та ймовірність знаходження глобального екстремуму; оператор репродукції замінений оператором клонування, а опе-

ратор схрещування не використовується, що підвищує швидкість пошуку розв'язку.

Як видно з рисунка 1, алгоритм селекції клонів починається з визначення цільової функції, яку необхідно оптимізувати. Створюються деякі можливі рішення-кандидати. Антитіла використовуються в цільовій функції для обчислення їх спорідненості. Це дозволяє визначити ті з них, які будуть клоновані для наступного кроку. Клоновані значення змінюються з попередньо визначеним співвідношенням, а спорідненість перераховується та сортується. Після проведення оцінки спорідненості та, що матиме найменше значення, і буде розв'язком, найближчим для вирішення поставленого завдання.

Мета та задачі дослідження. Метою роботи є підвищення ефективності кластеризації квазіперіодичного сигналу за рахунок синтезу методу перетворення такого сигналу й ієрархічно-ітеративного методу кластеризації на основі алгоритму клонального відбору.

Для досягнення поставленої мети необхідно вирішити такі завдання:

- розробити метод перетворення квазіперіодичного сигналу;
- запропонувати критерій якості кластеризації квазіперіодичного сигналу;
- створити ієрархічно-ітеративний метод кластеризації квазіперіодичного сигналу на основі алгоритму клонального відбору;
- провести чисельне дослідження запропонованого методу кластеризації.

1. Метод перетворення квазіперіодичного одновимірної сигналу. Включає наступні етапи:

- розбиття квазіперіодичного сигналу на дискретні зразки;
- зсув дискретних зразків за часом та амплітудою;
- інтерполяції дискретних зразків;
- зсув та масштабування неперервних зразків за часом;
- зсув та масштабування неперервних зразків за амплітудою;
- дискретизація неперервних зразків.

1.1. Розбиття квазіперіодичного сигналу на дискретні зразки. Визначимо кінцеву множину дискретних зразків квазіперіодичного сигналу, що описуються сімейством цілих обмежених фінітних дискретних функцій $X = \{x_i \mid i \in \{1, \dots, I\}\}$, у вигляді

$$x_i(n) = \begin{cases} f(n), & n \in \{N_i^{\min}, \dots, N_i^{\max}\} \\ 0, & n \notin \{N_i^{\min}, \dots, N_i^{\max}\} \end{cases}, \quad i \in \{1, \dots, I\},$$

$$A_i^{\min} = \min_n f(n), \quad n \in \{N_i^{\min}, \dots, N_i^{\max}\}, \quad i \in \{1, \dots, I\},$$

$$A_i^{\max} = \max_n f(n), \quad n \in \{N_i^{\min}, \dots, N_i^{\max}\}, \quad i \in \{1, \dots, I\},$$

де $A_i^{\min}, A_i^{\max}$ – мінімальне та максимальне значення функції x_i на відрізку $\{N_i^{\min}, \dots, N_i^{\max}\}$;

I – потужність множини зразків квазіперіодичного сигналу.

1.2. Зсув дискретних зразків за часом та амплітудою. Визначимо кінцеву множину зсунутих за часом та амплітудою дискретних зразків, що описуються кінцевим сімейством цілочисельних обмежених фінітних дискретних функцій $X^s = \{x_i^s \mid i \in \{1, \dots, I\}\}$, у вигляді

$$x_i^s(n) = \begin{cases} x_i(n + N_i^{\min}) - A_i^{\min}, & n \in \{0, \dots, N_i\} \\ 0, & n \notin \{0, \dots, N_i\} \end{cases},$$

$$i \in \{1, \dots, I\},$$

$$N_i = N_i^{\max} - N_i^{\min},$$

$$A_i = A_i^{\max} - A_i^{\min}.$$

1.3. Інтерполяції дискретних зразків.

У статті було обрано лінійну інтерполяцію, яка вимагає найменшої обчислювальної складності.

Визначимо кінцеву множину неперервних зразків, які отримані в результаті лінійної інтерполяції й описуються кінцевим сімейством дійсних обмежених фінітних неперервних функцій $\Psi = \{\psi_i \mid i \in \{1, \dots, I\}\}$, у вигляді

$$\forall t \in [-\Delta t, T_i + \Delta t], \quad \psi_i(t) =$$

$$= \sum_{n=-1}^{N_i} \chi_{(t_n, t_{n+1})}(t) \left(x_i^s(n) + \frac{x_i^s(n+1) - x_i^s(n)}{\Delta t} (t - t_n) \right) +$$

$$+ \sum_{n=-1}^{N_i+1} \chi_{\{t_n\}}(t) x_i^s(n), \quad i \in \{1, \dots, I\},$$

$$\forall t \notin [-\Delta t, T_i + \Delta t] \quad \psi_i(t) = 0,$$

$$T_i = N_i \Delta t,$$

$$t_n = n \Delta t,$$

$$\chi_B(t) = \begin{cases} 1, & t \in B; \\ 0, & t \notin B, \end{cases} \quad \text{– індикаторна функція;}$$

де Δt – крок квантування за часом.

1.4. Зсув і масштабування неперервних зразків за часом. Визначимо кінцеву множину зсунутих та масштабованих за часом неперервних зразків, що описуються кінцевим сімейством дійсних обмежених фінітних неперервних функцій $\Psi^s = \{\psi_i^s \mid i \in \{1, \dots, I\}\}$, у вигляді

$$\forall t \in [\tilde{T}^{\min}, \tilde{T}^{\max}], \psi_i^s(t) = \psi_i \left(T_i \frac{t - \tilde{T}^{\min}}{\tilde{T}^{\max} - \tilde{T}^{\min}} \right),$$

$$\forall t \notin [\tilde{T}^{\min}, \tilde{T}^{\max}], \psi_i^s(t) = 0,$$

де $[\tilde{T}^{\min}, \tilde{T}^{\max}]$ – заданий та єдиний для всіх зразків відрізок.

У статті пропонується визначити $\tilde{T}^{\min}, \tilde{T}^{\max}$ наступним чином:

$$\tilde{T}^{\min} = \Delta t,$$

$$\tilde{T}^{\max} = \left(1 + \frac{1}{I} \sum_{i=1}^I (N_i^{\max} - N_i^{\min}) \right) \Delta t.$$

1.5. Зсув і масштабування неперервних зразків за амплітудою. Визначимо кінцеву множину зсунутих та масштабованих за амплітудою неперервних зразків, що описуються кінцевим сімейством дійсних обмежених фінітних неперервних функцій $\Psi^{ss} = \{\psi_i^{ss} \mid i \in \{1, \dots, I\}\}$, у вигляді

$$\forall t \in [\tilde{T}^{\min}, \tilde{T}^{\max}],$$

$$\psi_i^{ss}(t) = \tilde{A}^{\min} + \frac{\tilde{A}^{\max} - \tilde{A}^{\min}}{\tilde{A}_i^{\max} - \tilde{A}_i^{\min}} \psi_i^s(t),$$

$$\forall t \notin [\tilde{T}^{\min}, \tilde{T}^{\max}], \psi_i^{ss}(t) = 0,$$

$$\tilde{A}_i^{\max} = \max_t \psi_i^s(t), \tilde{A}_i^{\min} = \min_t \psi_i^s(t),$$

$$t \in [\tilde{T}^{\min}, \tilde{T}^{\max}],$$

де $\tilde{A}_i^{\min}, \tilde{A}_i^{\max}$ – мінімальне та максимальне значення функції ψ_i^s на відрізку $[\tilde{T}^{\min}, \tilde{T}^{\max}]$;

$\tilde{A}^{\min}, \tilde{A}^{\max}$ – задане та єдине для всіх зразків мінімальне та максимальне значення на заданому відрізку $[\tilde{T}^{\min}, \tilde{T}^{\max}]$.

У статті пропонується визначити $\tilde{A}^{\min}, \tilde{A}^{\max}$ наступним чином:

$$\tilde{A}^{\min} = 0, \tilde{A}^{\max} = \frac{1}{I} \sum_{i=1}^I (A_i^{\max} - A_i^{\min}).$$

1.6. Дискретизація неперервних зразків. Визначимо кінцеву множину дискретних зразків, які отримані з неперервних шляхом

дискретизації за часом й описуються кінцевим сімейством цілочисельних обмежених фінітних дискретних функцій $S = \{s_i \mid i \in \{1, \dots, I\}\}$, у вигляді

$$s_i(n) = \text{round}(\psi_i^{ss}(n\Delta t)), n \in \{\tilde{N}^{\min}, \dots, \tilde{N}^{\max}\},$$

$$\tilde{N}^{\min} = \tilde{T}^{\min} / \Delta t,$$

$$\tilde{N}^{\max} = \tilde{T}^{\max} / \Delta t,$$

де $\text{round}()$ – функція, яка округляє число до найближчого цілого.

Таким чином, запропонований метод перетворення квазіперіодичного сигналу перетворює зразки (квазіперіодичні ділянки) цього сигналу до єдиного амплітудно-часового вікна, шляхом зсуву та масштабування за часом та амплітудою, інтерполяцією та дискретизацією, що дозволяє надалі порівняти зразки сигналу різної довжини та з різним розмахом амплітуд.

2. Критерій якості кластеризації квазіперіодичного сигналу. Для оцінювання методу кластеризації в роботі запропоновано критерій якості, що визначає вибір такого вектора h , який містить номери кластерів для кожного зразка та доставляє мінімум нормованій внутрішньокластерній відстані:

$$F = \frac{\sum_{i=1}^I \sum_{k=1}^K \chi_{S_k}(s_i) \|s_i - \mu_k\|^2}{(\tilde{N}^{\max} - \tilde{N}^{\min} + 1) \tilde{A}^{\max} \sum_{i=1}^I \sum_{k=1}^K \chi_{S_k}(s_i)} \rightarrow \min_h,$$

де S_k – k -й кластер;

μ_k – центр k -го кластера;

$\chi_{S_k}(s_i)$ – індикаторна функція, яка опи-

сує належність s_i кластеру S_k .

3. Ієрархічно-ітеративний метод кластеризації на основі алгоритму клонального відбору. Розглядатимемо сімейство S , отримане в результаті застосування методу перетворення квазіперіодичного сигналу, що знаходиться в єдиному амплітудно-часовому вікні, як множину об'єктів кластеризації.

Запропонований метод кластеризації включає наступні етапи:

0. Початкова кількість кластерів $K = 2$.

1. Створюється початкова популяція $H = \{h\}$, причому кожне антитіло цієї популяції представлено у вигляді $h = (h_1, \dots, h_I)$. Кожна компонента h_i взаємно однозначно

відповідає об'єкту кластеризації та містить номер кластера, що визначається у вигляді

$$h_i = \text{round}((K-1) \text{rand}() + 1),$$

де $\text{round}()$ – функція, що округлює число до найближчого цілого;

$\text{rand}()$ – функція, що генерує рівномірно розподілене випадкове число в діапазоні $[0, 1]$.

2. Для кожного антитіла h обчислюються індикаторні функції $\chi_{S_k}(s_i)$, які описують належність s_i кластеру S_k , $S_k \subseteq S$ (набувають 1, якщо належать, і 0 в іншому випадку):

$$\chi_{S_k}(s_i) = \begin{cases} 1, & h_i = k \\ 0, & h_i \neq k \end{cases}, i \in \{1, \dots, I\}, k \in \{1, \dots, K\},$$

при цьому повинні виконуватись умови:

$$\sum_{k=1}^K \chi_{S_k}(s_i) = 1, i \in \{1, \dots, I\},$$

$$\sum_{i=1}^I \chi_{S_k}(s_i) > 0, k \in \{1, \dots, K\},$$

$$\chi_{S_k}(s_i) \in \{0, 1\}, k \in \{1, \dots, K\}, i \in \{1, \dots, I\}.$$

3. Присвоюється номер ітерації $\tau = 1$.

4. Для кожного антитіла h обчислюються центроїди кластерів:

$$\mu_k(m) = \frac{\sum_{i=1}^I \chi_{S_k}(s_i) s_r(m)}{\sum_{i=1}^I \chi_{S_k}(s_i)}, k \in \{1, \dots, K\},$$

$$m \in \{\tilde{N}^{\min}, \dots, \tilde{N}^{\max}\}.$$

5. Для кожного антитіла h обчислюється значення фітнес-функції.

Фітнес-функція (цільова функція) визначена у вигляді

$$F(h) = \frac{\sum_{i=1}^I \sum_{k=1}^K \chi_{S_k}(s_i) \|s_i - \mu_k\|^2}{(\tilde{N}^{\max} - \tilde{N}^{\min} + 1) \tilde{A}^{\max} \sum_{i=1}^I \sum_{k=1}^K \chi_{S_k}(s_i)} \rightarrow \min_h,$$

$$\|s_i - \mu_k\| = \sqrt{\sum_{m=\tilde{N}^{\min}}^{\tilde{N}^{\max}} |s_i(m) - \mu_k(m)|}, i \in \{1, \dots, I\},$$

$$k \in \{1, \dots, K\}.$$

$$\hat{b}_q = \begin{cases} 1, & (\text{rand}() \geq p(h) \wedge (b_q = 1)) \vee ((\text{rand}() < p(h)) \wedge (b_q = 0)); \\ 0, & (\text{rand}() \geq p(h) \wedge (b_q = 0)) \vee ((\text{rand}() < p(h)) \wedge (b_q = 1)), \end{cases}$$

$$q \in \{1, \dots, I \cdot E\},$$

6. Для кожного антитіла h обчислюється афінність.

Афінність визначає близькість поточного антитіла до кращого антитіла та обчислюється на основі функції корисності у вигляді

$$\Phi(h) = 1 - \frac{F(h) - \min_{h \in H} F(h)}{\max_{h \in H} F(h) - \min_{h \in H} F(h)},$$

де $\Phi(h) \in [0, 1]$.

Якщо $\Phi(h) = 1$, то i -те антитіло є кращим. Якщо $\Phi(h) = 0$, то i -те антитіло є гіршим.

7. Для кожного антитіла h виконується оператор клонування.

Оператор клонування грає роль, аналогічну до оператора репродукції генетичного алгоритму.

Зазвичай кількість клонів q для кожного антитіла h визначається як

$$q = \text{round}(\alpha | H|),$$

де α – параметр клонування, $\alpha > 0$.

У результаті застосування оператора клонування формується множина клонів антитіла $Z = \{z\}$, причому кожний клон представлений у вигляді $z = (z_1, \dots, z_I)$.

8. Для кожного клону z виконується оператор мутації.

Оператор мутації дозволяє отримати з клонів антитіла нові антитіла з властивостями, що різко відрізняються за Евклідовою метрикою.

Перед мутацією кожний цілочисельний клон $z = (z_1, \dots, z_I)$ взаємно однозначно відображається в бінарний клон виду $b = (b_1, \dots, b_{I \cdot E})$ у вигляді

$$b_{iE+1-e} = \lceil 2^{1-e} \cdot z_i \rceil \bmod 2, i \in \overline{1, I}, e \in \overline{1, E},$$

де E – кількість біт для кодування однієї компоненти клону;

$\lceil \cdot \rceil$ – функція, що визначає цілу частину.

Мутація над кожним бінарним клоном b виконується у вигляді

$p(h) = e^{-\beta \Phi(h)}$ – ймовірність мутації антитіла h ,

де β – параметр мутації, $\beta > 0$. Чим більше β , тим менша ймовірність мутації.

Після завершення мутації виконується взаємно однозначне відображення кожного бінарного клону $\hat{b} = (\hat{b}_1, \dots, \hat{b}_{I-E})$ в цілочисельний клон $\hat{z} = (\hat{z}_1, \dots, \hat{z}_I)$ у вигляді

$$\hat{z}_i = \sum_{e=1}^E (2^{E-e} \cdot \hat{b}_{iE+1-e}), \quad i \in \overline{1, I}.$$

У результаті застосування оператора мутації формується множина клонів $\hat{Z} = \{\hat{z}\}$, що мутували.

9. Виконується оператор редукції.

Як оператор редукції використовується схема селекції, яка забезпечує спрямованість пошуку (найкращі антитіла зберігаються) і полягає в наступному. Антитіла попередньої популяції і клони, що мутували, об'єднуються в проміжну популяцію й впорядковуються за афінністю. У нову популяцію відбираються $|H|$ перших (найкращих за афінністю) особин проміжної популяції.

10. Виконується оператор заміни.

Для ширшого дослідження простору пошуку використовується заміна останніх (найгірших за афінністю) антитіл популяції новими антитілами. Кількість антитіл d , що замінюються, фіксована і становить від 5 % до 20 % розміру популяції.

Кожне нове антитіло представлене у вигляді $h = (h_1, \dots, h_I)$, а кожна його компонента генерується у вигляді

$$h_i = \text{round}((K-1)\text{rand}() + 1), \quad i \in \overline{1, I}.$$

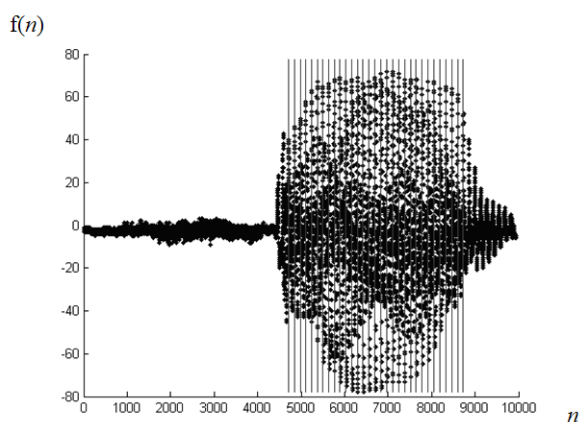


Рисунок 2 – Розбиття квазіперіодичного звуку «о» слова «со» на дискретні зразки

11. Визначається найкраще антитіло за фітнес-функцією

$$h^* = \arg \min_{h \in H} F(h).$$

12. Перевіряється умова завершення.

При $\frac{F(h^*)}{\bar{A}^{\max} \sqrt{\bar{N}^{\max}}} \leq \varepsilon$ отримано задовіль-

ний результат методу кластеризації h^* , де ε – задане значення точності кластеризації.

Якщо $\frac{F(h^*)}{\bar{A}^{\max} \sqrt{\bar{N}^{\max}}} > \varepsilon$ і $\tau < \tau^{\max}$, то

$\tau = \tau + 1$, перехід на 4.

Якщо $\frac{F(h^*)}{\bar{A}^{\max} \sqrt{\bar{N}^{\max}}} > \varepsilon$ і $\tau = \tau^{\max}$, то

$K = K + 1$, перехід на 1.

У результаті роботи методу кластеризації буде визначено структуру квазіперіодичного сигналу, що задається h^* .

Відтак, запропонований метод кластеризації квазіперіодичного сигналу на основі ієрархічно-ітеративного підходу та алгоритму клонального відбору зменшує чутливість до початкових значень центротидів кластерів за рахунок випадкового пошуку й забезпечує адаптивне налаштування кількості кластерів завдяки ієрархічному підходу.

Результати досліджень. Чисельне дослідження проводилося на основі квазіперіодичного звуку мовленнєвого сигналу на комп'ютері з процесором Intel Pentium Quad-Core з тактовою частотою 2.58 ГГц в середовищі MATLAB.

На рисунках 2-7 зображено приклади роботи методу перетворення квазіперіодичного одновимірної сигналу.

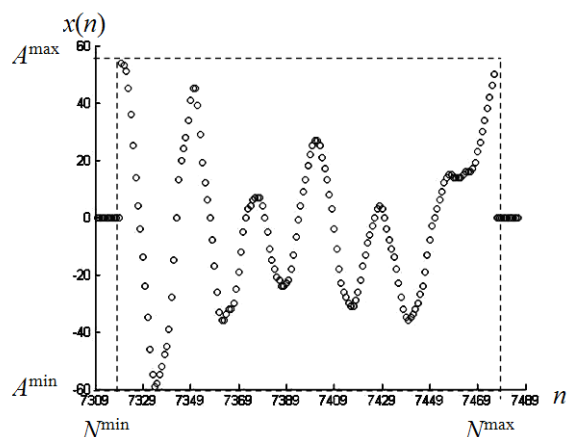


Рисунок 3 – Початковий дискретний зразок

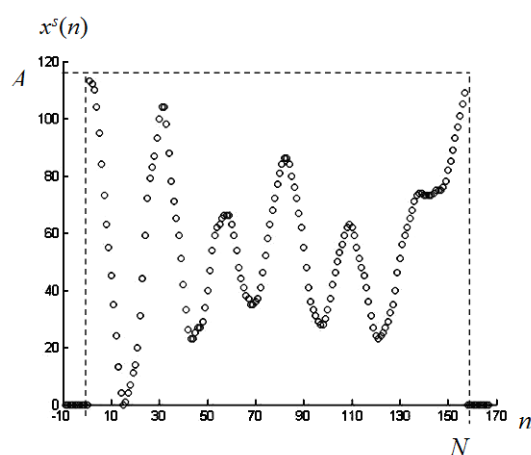


Рисунок 4 – Зсунутий за часом та амплітудою дискретний зразок

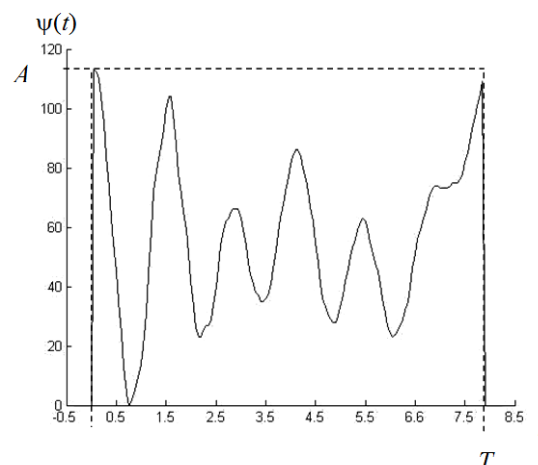


Рисунок 5 – Неперервний зразок

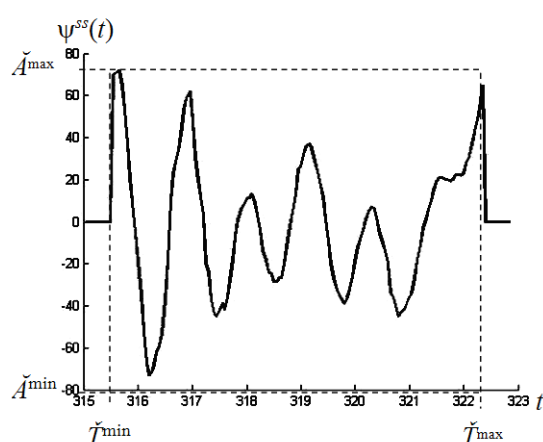


Рисунок 6 – Зсунутий і масштабований за часом та амплітудою неперервний зразок

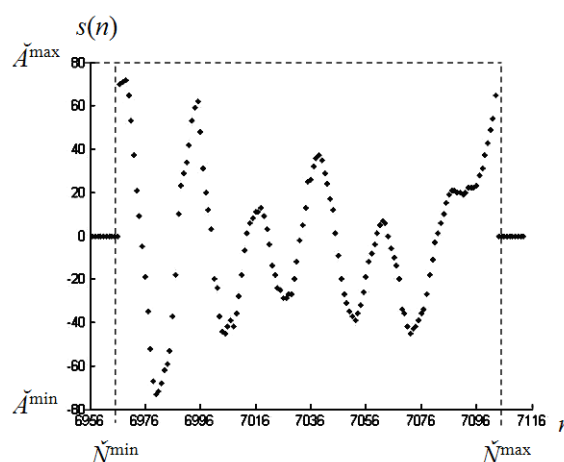


Рисунок 7 – Результуючий дискретний зразок

На рисунку 2 зображено розбиття квазі-періодичного звуку «о» слова «со» на дискретні зразки. Наведений ансамбль підтверджує складність візуального сприйняття акустичної характеристики зразка. Тому для усунення цього недоліку з дискретних зразків вибирається один, який і буде підданий перетворенню.

На рисунку 3 зображено початковий дискретний зразок, що знаходиться на відрізку $\{N_i^{\min}, \dots, N_i^{\max}\}$ з мінімальним та максимальним значенням амплітуди $A_i^{\min}, A_i^{\max}$ і піддається подальшим перетворенням, показаним на рисунках 3-6.

На рисунку 4 зображено дискретний зразок після зсуву за часом та амплітудою, що знаходиться на відрізку $\{0, \dots, N\}$ з мінімальним та максимальним значенням $0, A$.

На рисунку 5 зображено неперервний зразок сигналу, отриманий з дискретного зразка в результаті лінійної інтерполяції з частотою 22050 Гц та кроком дискретизації за часом $\Delta t = 1/22050$, що знаходиться на відрізку $[0, T]$ з мінімальним та максимальним значенням $0, A$.

На рисунку 6 зображено неперервний зразок сигналу після зсуву та масштабування за часом та амплітудою, що знаходиться на відрізку $[\tilde{T}^{\min}, \tilde{T}^{\max}]$ з мінімальним та максимальним значенням $\tilde{A}^{\min}, \tilde{A}^{\max}$.

На рисунку 7 зображено результуючий дискретний зразок сигналу, отриманий з неперервного зразка в результаті дискретизації з частотою 22050 Гц та кроком дискретизації $\Delta t = 1/22050$. Перетворений зразок знаходиться на відрізку $\{\tilde{N}^{\min}, \dots, \tilde{N}^{\max}\}$ з мінімальним та максимальним значенням $\tilde{A}^{\min}, \tilde{A}^{\max}$.

У таблиці 1 наведено порівняння запропонованого авторами методу та існуючих методів.

тодів кластеризації, при цьому параметр клонування $\alpha = 0.1$, параметр мутації $\beta = 2.5$, кількість антитіл $d = 0.2 |H|$, що замінюються.

Таблиця 1 – Оцінка методів кластеризації

Метод	Ймовірність кластеризації (%)
центроїдний (к-середніх)	0.80
центроїдний (нечітких с-середніх)	0.83
модель суміші розподілів (ЕМ)	0.85
коннекціоністський (мережа SOM)	0.82
медоїдний (субтрактивна кластеризація)	0.90
коннекціоністський (мережа ART)	0.93
авторський	0.98

Під ймовірністю кластеризації малося на увазі відношення правильно класифікованих зразків квазіперіодичного сигналу до їх загальної кількості. Кластеризація проводилася на квазіперіодичних звуках мовлення, вимовлених різними дикторами.

Обговорення результатів. Результат, наведений у таблиці 1, показує, що авторський метод забезпечує високу ймовірність кластеризації (використовує попередню обробку та адаптивне налаштування кількості кластерів).

На основі проведеного експерименту можна зробити висновок, що центроїдні методи кластеризації (к-середніх, нечітких с-середніх), метод моделі суміші розподілів (ЕМ) і коннекціоністський метод (мережа SOM) мають найменшу ймовірність кластеризації, оскільки мають фіксовану кількість кластерів.

Коннекціоністський метод (мережа ART) та медоїдний метод (субтрактивна кластеризація) мають більш високу ймовірність кластеризації, оскільки мають адаптивне налаштування кількості кластерів.

Візуалізація результатів (див. рисунки 2-7) застосування методу перетворення квазіперіодичного сигналу і чисельного дослідження ієрархічно-ітеративного методу кластеризації квазіперіодичного сигналу на основі алгоритму клонального відбору підтверджує ефективність запропонованого

методу для аналізу структури квазіперіодичного сигналу.

Висновки. У статті в межах створення методу перетворення та кластеризації квазіперіодичного сигналу було вирішено такі завдання:

– розроблено метод перетворення квазіперіодичного сигналу;

– створено ієрархічно-ітеративний метод кластеризації квазіперіодичного сигналу на основі алгоритму клонального відбору;

– проведено чисельне дослідження запропонованого методу кластеризації.

Наукова новизна проведеного дослідження полягає в наступному:

1. Удосконалено метод перетворення квазіперіодичного сигналу, який перетворює зразки (квазіперіодичні ділянки) цього сигналу до єдиного амплітудно-часового вікна, за допомогою зсуву та масштабування за часом і амплітудою, інтерполяції та дискретизації, що дозволяє надалі порівнювати зразки сигналу різної довжини і з різним розмахом амплітуд.

2. Отримав подальший розвиток метод кластеризації квазіперіодичного сигналу на основі ієрархічно-ітеративного підходу та алгоритму клонального відбору, який зменшує чутливість до початкових значень центроїдів кластерів за рахунок випадкового пошуку та забезпечує адаптивне налаштування кількості кластерів за рахунок ієрархічного підходу. Це підвищує можливість кластеризації до 0.98.

3. Отримав подальший розвиток критерій якості кластеризації квазіперіодичного сигналу, що дозволяє отримати нормовану внутрішньокластерну відстань, що, в свою чергу, дозволяє встановити єдиний поріг для оцінювання якості кластеризації.

Практичне значення. Запропонований метод дозволяє ставити і вирішувати завдання перетворення та кластеризації мовленнєвого або діагностичного сигналу, а також мережевого трафіку, що використовується для аналізу та зберігання біометричної, діагностичної та мережевої інформації в інтелектуальних комп'ютерних системах.

Перспективи подальших досліджень. Надалі передбачається досліджувати питання впливу методів перетворення та кластеризації на виявлення та класифікацію періодів у одновимірних квазіперіодичних сигналах.

Список використаних джерел

- [1] K. S. Thyagarajan, *Introduction to Digital Signal Processing using MATLAB with Application to Digital Communications*. Cham, Switzerland: Springer, 2020.
- [2] P. R. Hill, *Audio and Speech Processing with MATLAB*. Boca Raton, FL, USA: CRC Press, 2019.
- [3] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge, GB: Cambridge University Press, 2013.
- [4] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. New York, USA: Springer, 2013.
doi: 10.1007/978-1-4614-7138-7.
- [5] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity. The Lasso and Generalizations*. New York, USA: Chapman and Hall/CRC, 2015.
doi: 10.1201/b18401.
- [6] А. Н. Ткаченко, О. Ф. Грийо Тукало, О. В. Дзись, и С. М. Лаховец, "Метод кластеризации на основе последовательного запуска k-средних с усовершенствованным выбором кандидата на новую позицию вставки", *Наукові праці ВНТУ*, № 2, с. 1-10, 2012.
- [7] M. J. Brusco, E. Shireman, and D. Steinley, "A comparison of latent class, k-means, and k-median methods for clustering dichotomous data", *Psychological Methods*, 22 (3), pp. 563-580, 2017.
doi: 10.1037/met0000095.
- [8] C. C. Aggarwal, and C. K. Reddy, *Data Clustering: Algorithms and Applications*. Boca Raton, FL, USA: CRC Press, 2014.
- [9] Z. Fu, and L. Wang, "Color image segmentation using Gaussian mixture model and EM algorithm", in *Proc. Multimedia and Signal Processing. Communications in Computer and Information Science (CMSP'2012)*, 2012, vol. 346, pp. 61-66.
doi: 10.1007/978-3-642-35286-7_9.
- [10] T. Kohonen, *Self-Organizing and Associative Memory*, 3rd ed. Berlin, Heidelberg, Germany: Springer, 2012.
doi: 10.1007/978-3-642-88163-3.
- [11] T. Kohonen, "Essentials of the self-organizing map", *Neural Networks: The Official Journal of the Inter-National Neural Network Society*, vol. 37, pp. 52-65, 2013.
doi: 10.1016/j.neunet.2012.09.018.
- [12] K.-L. Du, and M. N. S. Swamy, *Neural Networks and Statistical Learning*. London, GB: Springer-Verlag, 2014.
doi: 10.1007/978-1-4471-5571-3.
- [13] Ю. А. Скобцов, и Е. Е. Федоров, *Метаэвристики*. Донецк, Украина: Ноулидж (Донецкое отд.), 2013.
- [14] J. Radosavljević, *Metaheuristic Optimization in Power Engineering*. New York, USA: Institution of Engineering and Technology, 2018.
doi: 10.1049/pbpo131e.
- [15] E. Alba, A. Nakib, and P. Siarry, *Metaheuristics for Dynamic Optimization*. Berlin, Germany: Springer-Verlag, 2013.
doi: 10.1007/978-3-642-30665-5.
- [16] K.-L. Du, and M. N. S. Swamy, *Search and Optimization by Metaheuristics. Techniques and Algorithms Inspired by Nature*. Charm, Switzerland: Springer, 2016.
doi: 10.1007/978-3-319-41192-7.
- [17] O. Bozorg Haddad, M. Solgi, and H. Loaiciga, *Meta-heuristic and Evolutionary Algorithms for Engineering Optimization*. Hoboken, New Jersey, USA: Wiley & Sons, 2017.
doi: 10.1002/9781119387053.
- [18] X.-S. Yang, *Nature-Inspired Algorithms and Applied Optimization*. Charm, Switzerland: Springer, 2018.
doi: 10.1007/978-3-319-67669-2.
- [19] L. N. De Castro, and F. J. Von Zuben, "Learning and optimization using the clonal selection principle", *IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems*, vol. 6, no. 3, pp. 239-251, 2002.
- [20] E. D. Ülker, and S. Ülker, "Comparison study for clonal selection algorithm and genetic algorithm", *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 4, no. 4, pp. 107-118, 2012.
doi: 10.5121/ijcsit.2012.4410

References

- [1] K. S. Thyagarajan, *Introduction to Digital Signal Processing using MATLAB with Application to Digital Communications*. Cham, Switzerland: Springer, 2020.
- [2] P. R. Hill, *Audio and Speech Processing with MATLAB*. Boca Raton, FL, USA: CRC Press, 2019.

- [3] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge, GB: Cambridge University Press, 2013.
- [4] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. New York, USA: Springer, 2013.
doi: 10.1007/978-1-4614-7138-7.
- [5] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity. The Lasso and Generalizations*. New York, USA: Chapman and Hall/CRC, 2015.
doi: 10.1201/b18401.
- [6] A. N. Tkachenko, O. F. Griyo Tukalo, O. V. Dzisy, and S. M. Lakhovets, "Clustering method based on sequential launch of k-means with improved selection of a candidate for a new insertion position", *Naukovi pratsi VNTU*, no. 2, pp. 1-10, 2012 [in Russian].
- [7] M. J. Brusco, E. Shireman, and D. Steinley, "A comparison of latent class, k-means, and k-median methods for clustering dichotomous data", *Psychological Methods*, 22 (3), pp. 563-580, 2017.
doi: 10.1037/met0000095.
- [8] C. C. Aggarwal, and C. K. Reddy, *Data Clustering: Algorithms and Applications*. Boca Raton, FL, USA: CRC Press, 2014.
- [9] Z. Fu, and L. Wang, "Color image segmentation using Gaussian mixture model and EM algorithm", in *Proc. Multimedia and Signal Processing. Communications in Computer and Information Science (CMSP 2012)*, 2012, vol. 346, pp. 61-66.
doi: 10.1007/978-3-642-35286-7_9
- [10] T. Kohonen, *Self-Organizing and Associative Memory*, 3rd ed. Berlin, Heidelberg, Germany: Springer, 2012.
doi: 10.1007/978-3-642-88163-3
- [11] T. Kohonen, "Essentials of the self-organizing map", *Neural Networks: The Official Journal of the Inter-National Neural Network Society*, vol. 37, pp. 52-65, 2013.
doi: 10.1016/j.neunet.2012.09.018.
- [12] K.-L. Du, and M. N. S. Swamy, *Neural Networks and Statistical Learning*. London, GB: Springer-Verlag, 2014.
doi: 10.1007/978-1-4471-5571-3.
- [13] Yu. A. Skobtsov, and E. E. Fedorov, *Metaheuristics*. Donetsk, Ukraine: Noulidzh (Donetsk branch), 2013 [in Russian].
- [14] J. Radosavljević, *Metaheuristic Optimization in Power Engineering*. New York, USA: Institution of Engineering and Technology, 2018.
doi: 10.1049/pbpo131e.
- [15] E. Alba, A. Nakib, and P. Siarry, *Metaheuristics for Dynamic Optimization*. Berlin, Germany: Springer-Verlag, 2013.
doi: 10.1007/978-3-642-30665-5.
- [16] K.-L. Du, and M. N. S. Swamy, *Search and Optimization by Metaheuristics. Techniques and Algorithms Inspired by Nature*. Charm, Switzerland: Springer, 2016.
doi: 10.1007/978-3-319-41192-7.
- [17] O. Bozorg Haddad, M. Solgi, and H. Loaiciga, *Meta-heuristic and Evolutionary Algorithms for Engineering Optimization*. Hoboken, New Jersey, USA: Wiley & Sons, 2017.
doi: 10.1002/9781119387053.
- [18] X.-S. Yang, *Nature-Inspired Algorithms and Applied Optimization*. Charm, Switzerland: Springer, 2018.
doi: 10.1007/978-3-319-67669-2.
- [19] L. N. De Castro, and F. J. Von Zuben, "Learning and optimization using the clonal selection principle", *IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems*, vol. 6, no. 3, pp. 239-251, 2002.
- [20] E. D. Ülker, and S. Ülker, "Comparison study for clonal selection algorithm and genetic algorithm", *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 4, no. 4, pp. 107-118, 2012.
doi: 10.5121/ijcsit.2012.4410.

E. E. Fedorov, Dr. Tech. Sc., Professor,
T. Yu. Utkina, Ph. D., Associate Professor
e-mail: t.utkina@chdtu.edu.ua
Cherkasy State Technological University
Shevchenko blvd, 460, Cherkasy, 18006, Ukraine

METHOD OF CLUSTERIZATION OF QUASIPERIODIC SIGNAL BASED ON CLONAL SELECTION ALGORITHM

The paper proposes a hierarchical-iterative method of clustering of a quasi-periodic signal based on clonal selection algorithm, which increases the speed and accuracy of clustering. Preliminary transformation of samples (quasi-periodic areas) of this signal into a single amplitude-time window based on shift and scaling in time and amplitude, linear interpolation and time sampling is performed.

Signals processed by intelligent computer systems for identification, analysis and synthesis, storage of digital signals (such as acoustic, graphic, vibration, electrogram, communication ones, etc.) are often quasi-periodic. This raises the problem of constructing effective methods for analyzing the structure of a quasi-periodic signal. One of the means of analyzing the structure of a quasi-periodic signal is clustering, which is a type of machine learning without a teacher.

As a result of the analysis of modern methods of clustering of quasi-periodic signals, it has been found that most of them have one or more of the following disadvantages: unknown exact number of clusters; sensitivity to initial values of the centroids of clusters; low probability of clustering; low speed of clustering; comparison of signal areas that have only the same size; comparison of only binary signals.

Therefore, it is important to develop a method for transforming a quasi-periodic signal and a hierarchical-iterative method of clustering based on a clonal selection algorithm. This will increase the efficiency for the analysis of quasi-periodic signal structure in digital data processing in intelligent computer systems of identity identification, technical and medical diagnostics, network traffic analysis, etc.

A comparison of the method proposed by the authors and existing clustering methods is given, with the cloning parameter $\alpha = 0.1$, the mutation parameter $\beta = 2.5$, the number of replacement antibodies $d = 0.2 \cdot H$. Clustering has been performed on quasi-periodic speech sounds uttered by different speakers. A method for converting a quasi-periodic signal, which converts samples (quasi-periodic areas) of this signal into a single amplitude-time window by shifting and scaling in time and amplitude, interpolation and sampling, is proposed. This allows to compare signal samples of different lengths and with different amplitudes. A method for clustering a quasi-periodic signal based on a hierarchical-iterative approach and a clonal selection algorithm has been developed, which reduces the sensitivity to initial values of cluster centroids due to random search and provides adaptive adjustment of the cluster number due to the hierarchical approach, and also increases the probability of clustering to 0.98.

Keywords: quasi-periodic signal, clonal selection algorithm, hierarchical-iterative clustering, signal transformation, signal structure analysis.