DOI: 10.18372/2310-5461.67.20244
УДК 621.391

О. Ю. Лавриненко, канд. техн. наук, доц.
Державний університет "Київський авіаційний інститут"
orcid.org/0000-0002-7738-161X
e-mail: oleksandrlavrynenko@gmail.com

МЕТОД ОПТИМІЗАЦІЇ БАЗОВОЇ ВЕЙВЛЕТ-ФУНКЦІЇ ЗА КРИТЕРІЄМ МІНІМІЗАЦІЇ ПОМИЛКИ АПРОКСИМАЦІЇ МОВНИХ СИГНАЛІВ

Вступ

Однією з актуальних задач обробки мовних сигналів є їхня класифікація, перший етап якої - виділення ознак розпізнавання мови. Для цього, як правило, використовуються різні методи на базі спектрального аналізу. Найбільш широко застосовуються мел-частотні кепстральні коефіцієнти (МЧКК) [1], що розраховуються на основі перетворення Фур'є і дають змогу отримати ознаки розпізнавання мовних сигналів із врахуванням сприйняття звуку людиною. Однак цей метод не забезпечує достатньо високої роздільної здатності отриманих ознак розпізнавання, що призводить до зниження точності класифікації загалом і необхідності використання нелінійних методів класифікації.

Для аналізу нестаціонарних сигналів, до яких відносяться і мовні сигнали, широко застосовується вейвлет-перетворення, яке дає змогу отримати оптимальну частотно-часову роздільну здатність, що в свою чергу, дає можливість локалізувати малі зміни мовного сигналу як у частотній, так і в часовій областях. Ця властивість вейвлет-функцій відіграє значущу роль під час виділення ознак і класифікації мовних сигналів, з цієї причини було зроблено вибір – в якості базиса спектрального перетворення використовувати вейвлет-базис, а не базис Фур'є. Однак використання вейвлет-перетворення з різними базовими вейвлет-функціями для виділення вейвлет-ознак розпізнавання мовних сигналів [2] не дає змоги досягти точності класифікації, істотно вищою, ніж у разі використання МЧКК на основі базису Фур'є.

Таким чином, виходячи з вищенаведеного, ще одним із найактуальніших завдань у галузі обробки та розпізнавання мовних сигналів є поліпшення роздільної здатності вейвлет-ознак розпізнавання мови.

Аналіз останніх досліджень та публікацій

Одним із підходів, який вирішує вищезазначену проблему є оптимізація параметрів побудови вейвлет-ознак розпізнавання мовних сигналів для розв'язуваної задачі класифікації [3]. Оптимізацію параметрів алгоритму виділення адаптивних вейвлет-ознак розпізнавання мовних сигналів можна здійснити з використанням чисельних методів оптимізації, зокрема широко використовується генетичний алгоритм.

Альтернативний підхід, який дає змогу поліпшити вейвлет-ознаки розпізнавання мови – агрегація ознак [4], отриманих із застосуванням різних підходів, однак це призводить до суттєвого збільшення розмірності ознакового простору, унаслідок чого простір стає розрідженим, що в свою чергу, погіршує точність подальшого статистичного аналізу та розпізнавання. Щоб уникнути збільшення ознакового простору, проводять редукцію ознакового простору [5], виходячи з припущення про те, що багато ознак розпізнавання мовних сигналів мають невисоку інформаційну цінність.

Такий підхід має деякі очевидні недоліки:

1) втрата інформації про мовний сигнал, яка може бути значущою для подальшого аналізу;

2) ознаки розпізнавання мови можуть бути розподілені в просторі ознак із великим перекриттям кластерів схожих мовних сигналів, що вимагає використання складних нелінійних класифікаторів.

У зв'язку з цим авторами пропонується метод побудови оптимальної базової вейвлет-функції за критерієм мінімізації помилки апроксимації для виділення адаптивних вейвлет-ознак розпізнавання мовних сигналів. Це дасть змогу забезпечити достатньо високу роздільну здатність отриманих вейвлет-ознак розпізнавання, що призведе до підвищення точності класифікації акустично схожих фонем мови.

Постановка задачі дослідження

Математичну формалізацію поставленої наукової задачі можна представити таким чином.

Шукану базову вейвлет-функцію описуватимемо за допомогою набору параметрів $\vec{x} = (x_1, \ldots, x_n)$, на які накладаються обмеження виду $X = \{\vec{x} \mid 1 \geq x_i \geq 0, i = 1, \ldots, n\} \subset R^n$. Тоді часо-

© О. Ю. Лавриненко, 2025

ве представлення шуканої функцій обчислюватимемо за допомогою сплайну Акіми на основі даного набору параметрів, який являє собою ординати базових точок вейвлет-функції [6].

Для стійкої класифікації фонем шукана базова вейвлет-функція має локалізувати значущі коефіцієнти для фонем зі схожим механізмом утворення в схожій смузі масштабів. Зробимо апріорне припущення про те, що в даному випадку для моменту часу t значення вейвлет-коефіцієнтів у смузі масштабів

$$S = [s_c - \delta, s_c + \delta],$$

де s_c – центральне значення масштабів даної смуги, можна апроксимувати за допомогою функції Гауса [7]

$$G_t(s) = \frac{1}{\sigma_t \sqrt{2\pi}} \exp\left(-\frac{(s - s_c)^2}{2\sigma_t{}^2} \right),$$

де σ_t – дисперсія вейвлет коефіцієнтів у момент часу t, при цьому ширина смуги $2\delta < 3\sigma$.

Тоді цільову функцію оптимізації [8] задамо як величину, обернену сумарній помилці апроксимації

$$F(\vec{x}) = \left(\sum_{t=0}^{N} \sum_{s=s_l}^{s_h} \left| G_t(s) - W_{\vec{x}}(s,t) \right| \right)^{-1},$$

і будемо шукати її максимум. Тут $W_{\vec{x}}(s,t)$ – коефіцієнти вейвлет-перетворення, заданого за допомогою базової вейвлет-функції, отриманої з набору параметрів $\vec{x}$.

Виклад основного матеріалу

Блок-схему розробленого методу виділення адаптивних вейвлет-ознак розпізнавання мовних сигналів на основі оптимізації базової вейвлет-функції за критерієм мінімізації помилки апроксимації представлено на рис. 1.

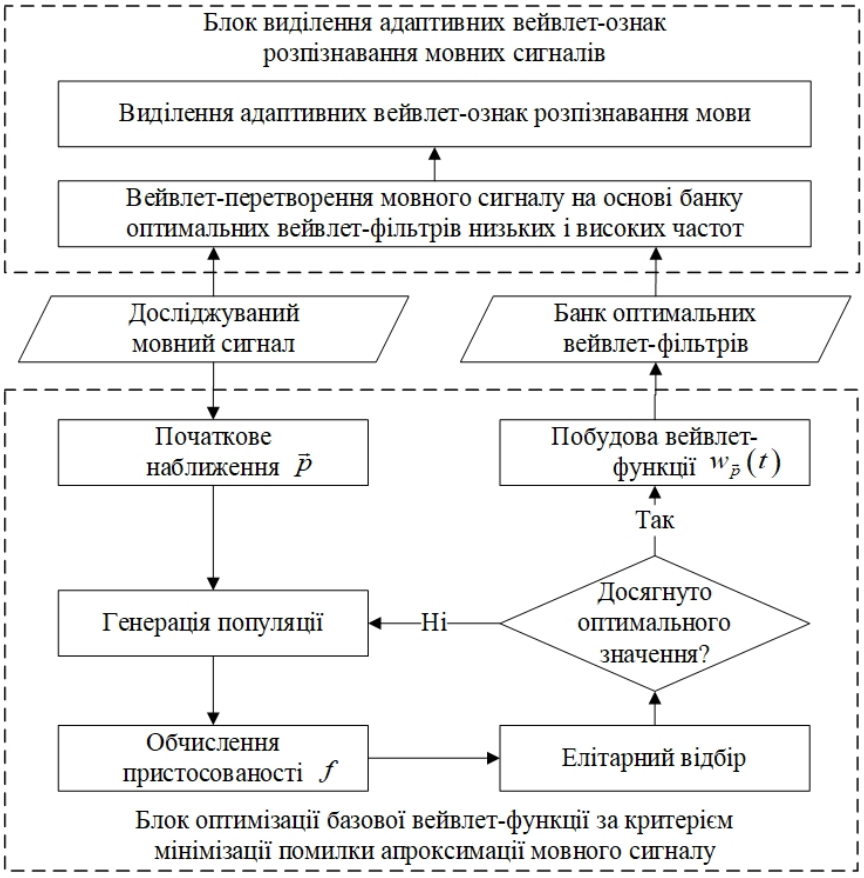

Рис. 1. Блок-схема методу виділення адаптивних вейвлет-ознак розпізнавання мовних сигналів на основі оптимізації базової вейвлет-функції за критерієм мінімізації помилки апроксимації

Вейвлет-перетворення являє собою інтегральне перетворення, яке задається функцією двох змінних:

$$W(\alpha,\tau) = \int_{-\infty}^{+\infty} \frac{s(t)}{\sqrt{\alpha}} w\left(\frac{t-\tau}{\alpha} \right) dt, \qquad (1)$$

де $s(t)$ – аналізований мовний сигнал, $w(t)$ – вейвлет-функція, α – масштаб, τ – зсув у часі.

Припустімо, що існує вейвлет-функція $w(t)$, форму якої задано набором параметрів $\vec{p}$ і яка дає змогу локалізувати коефіцієнти вейвлет-перетворення $W(\alpha,\tau)$ у такий спосіб, щоб забез-

© О. Ю. Лавриненко, 2025

печити найкращий поділ заданих класів у ознаковому просторі розпізнавання мови [9].

Для оцінки розподілу вейвлет-ознак у ознаковому просторі будемо використовувати лінійний класифікатор на базі методу опорних векторів. Для випадку бінарної класифікації цей метод дає змогу побудувати оптимальну розділювальну гіперплощину $\langle \vec{b}, \vec{x} \rangle - b_0 = 0$, максимізуючи водночас відстань між розділювальною гіперплощиною і межами класів. Вирішальне правило для такого класифікатора описується виразом

$$a(\vec{x}) = sign\left(\langle \vec{b}, \vec{x} \rangle - b_0\right). \qquad (2)$$

З використанням вейвлет-перетворення (1) побудуємо набір вейвлет-ознак $X = \{\vec{x}_i \mid i = 1,\ldots,N\}$ для мовних сигналів двох різних класів, точна класифікація яких відома $c(\vec{x}) = \{-1;1\}$.

Далі, застосовуючи процедуру крос-валідації, побудуємо вирішальне правило за допомогою (2). Для цього розділимо множину X на K підмножин, що не перетинаються

$$\left\{ \begin{matrix} X_i \mid X = X_1 \cup X_2 \cup,\ldots,\cup X_k; X_i \cap X_j = \varnothing; \\ i = 1,\ldots,K; j = 1,\ldots,K; i \neq j \end{matrix} \right\},$$

при цьому кожна підмножина X_i містить n_k векторів вейвлет-ознак [10].

Навчимо набір класифікаторів

$$\{a_k(\vec{x}) \mid k = 1,\ldots,K\}, \qquad (3)$$

використовуючи набір векторів вейвлет-ознак $X^l = X / X_k$. Як цільову функцію використовуватимемо середню точність класифікації набором класифікаторів (3), яка може бути задана виразом:

$$f = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{n_k}\sum_{\vec{x}\in X_k}\frac{\left|a_k(\vec{x}) + c(\vec{x})\right|}{2}, \qquad (4)$$

де $a_k(\vec{x})$ – результат класифікації, $c(\vec{x})$ – правильний клас для даного вектора вейвлет-ознак, $a_k(\vec{x})$ і $c(\vec{x})$ можуть набувати тільки значення ± 1. Для побудови необхідної адаптивної вейвлет-функції необхідно знайти вектор параметрів $\vec{p} = (p_1, p_2, \ldots, p_n)$, $p_i \in R$, що задає вейвлет-перетворення $W_f(\alpha, \tau)$ такий, що $f \to \max$. Для побудови функції вейвлет-перетворення, адаптованої для класифікації досліджуваних мовних сигналів, використовувались чисельні методи оптимізації [11].

В якості метода оптимізації цільової функції в цій роботі було обрано генетичний алгоритм, оскільки він має низку переваг, найважливішими з яких є: 1) можливість знаходження глобального екстремуму цільової функції; 2) швидкість сходження; 3) можливість розпаралелювання алгоритму, що дає змогу суттєво скоротити час розрахунку за використання сучасних систем паралельного обробки даних.

Генетичний алгоритм оперує вектором $\vec{p}$ опосередковано через послідовність кодових символів $\vec{q} = (q_1, q_2, \ldots, q_n)$, яку в теорії еволюційних методів оптимізації прийнято називати хромосомою. Хромосома $\vec{q}$ однозначно визначає вектор параметрів $\vec{p}$, при цьому: 1) кожен параметр p_i, $i \in 1,\ldots,N$ описується відповідним геном q_i, $i \in 1,\ldots,N$; 2) кожний ген q_i, $i \in 1,\ldots,N$ складається з M алелів, які обирають зі скінченної множини. Для зручності реалізації генетичного алгоритму використовується скінченна множина алелів $\{0,1\}$.

Для побудови адаптивної функції вейвлет-перетворення часове представлення шуканої функції задавали параметричною кривою, а саме сплайном Акіми [12]. Такий вибір параметричної кривої зумовлений, по-перше, тим фактом, що результуюча крива проходить через усі точки, по-друге, стійкістю сплайна Акіми до локальних викидів – у цього сплайна практично відсутні осциляції кривої поблизу точок викиду, на відміну від кубічних сплайнів. Ця властивість сплайна Акіми є значущою, оскільки будь-які додаткові осциляції погіршують локалізацію вейвлет-функції в частотній області.

Сплайн задається набором ординат базових точок $p(n) = \{p_i \mid i = 1,\ldots,N\}$, де значення p_i кодується геном q_i, причому $p_i \in [-1,1]$. Оскільки ген складається з кінцевого числа алелів, відповідно значення кодуються з деякою точністю ε. У такому разі максимальна кількість значень, які можуть кодуватися геном q_i, описується виразом

$$K_{\max} = \frac{\max(p_i) - \min(p_i)}{\varepsilon},$$

а мінімальна кількість алелів у гені, необхідна для кодування всіх значень, становить

$$L(q_i) = \left\lceil \log_2\left(\frac{\max(p_i) - \min(p_i)}{\varepsilon}\right)\right\rceil.$$

Операції мутації та кросовера в даному випадку є тривіальними, оскільки будь-яка комбінація алелів є легітимною.

В якості алгоритму відбору осіб для формування наступної популяції використовували алгоритм елітарного відбору, оскільки цей алго-

© О. Ю. Лавриненко, 2025

ритм забезпечує вищу швидкість сходження під час розв'язання задачі оптимізації вейвлет-функції, в порівнянні з методами рулеточного або турнірного відборів.

Розв'язанням сформульованої вище задачі оптимізації є вектор параметрів $\vec{p}$, який описує деяку адаптивну вейвлет-функцію $w_p(t)$, що дає змогу локалізувати вейвлет-коефіцієнти в заданій смузі масштабів у такий спосіб, що функція $f \to \max$, що дає змогу підвищити роздільну здатність вейвлет-ознак розпізнавання мови.

Результати дослідження

За допомогою запропонованого методу побудовано базові оптимальні вейвлет-функції за критерієм мінімізації помилки апроксимації мовних сигналів, адаптовані для виділення вейвлет-ознак розпізнавання мовних сигналів, що використовуються в задачах аналізу та класифікації різних класів фонем мови. Для побудови адапти-вної вейвлет-функції застосовували описаний у даній статті метод із використанням генетичного алгоритму, при цьому кількість параметрів, що оптимізуються – $N = 250$, розмір популяції – 25000, $K_{\max} = 2000$, $L(q_i) = 11$.

Відповідно до графіка, наведеного на рис. 2, кількість ітерацій, необхідних для отримання адаптованої вейвлет-функції, становить 150. Для оцінки роздільної здатності адаптивних вей-влет-ознак розпізнавання мови, отриманих з використанням запропонованого методу, проведено порівняльне тестування різних ознак з різною розмірністю *d*. Тут порівнюється точність класифікації фонем бінарним класифікатором на основі методу опорних векторів з гаусовим ядром.

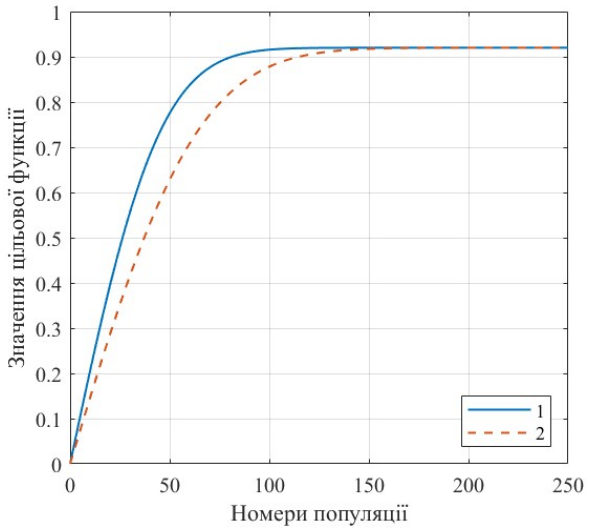

Рис. 2. Графік залежності значень цільової функції (4) від номера популяції: *1* – максимальне значення для поточної популяції; *2* – середнє значення

Для цього обрано ознаки, отримані з використанням вейвлет-перетворення на основі функцій Добеші $d = 13$, Мейєра $d = 13$; адаптивні вейвлет-ознаки (АВО) на основі запропонованого методу ($d_1 = 13$, $d_2 = 26$); ознаки, отримані з використанням МЧКК $d = 13$, дельта-МЧКК $d = 26$ (13 МЧКК + 13 дельта-коефіцієнтів), дельта-дельта-МЧКК $d = 39$ (13 МЧКК, 13 дельта-коефіцієнтів, 13 дельта-дельта-коефіцієнтів).

У табл. 1 подано результати лінійної бінарної класифікації фонем різних типів (голосні, сонорні, шумові). При цьому форма запису $[ph]$ – $[ph_1, ph_2, \ldots, ph_n]$ позначає розв'язання задачі класифікації фонеми $[ph]$ з множини фонем $[ph_1, ph_2, \ldots, ph_n]$.

Таблиця 1

Точність класифікації фонем української мови різних типів, %

Групи фонем	Добеши, $d = 13$	Мейєра, $d = 13$	АВО, $d = 13$	АВО, $d = 26$	МЧКК, $d = 13$	МЧКК, $d = 26$	МЧКК, $d = 39$
[а] – [а, о]	85,2	87,2	92,2	**95,4**	90,1	91,8	91,2
[а] – [голосні]	72,1	76,0	91,3	**94,7**	88,3	89,7	88,4
[а] – [всі фонеми]	71,4	74,9	85,6	**89,2**	84,6	82,4	81,8
[н] – [н, м]	58,4	63,7	75,4	**80,1**	66,9	70,3	69,9
[н] – [всі фонеми]	55,7	61,1	73,1	**78,3**	65,4	67,5	65,0
[з] – [з, ж, ш, с]	71,3	75,6	79,7	**85,7**	74,5	77,3	77,1
[з] – [всі фонеми]	68,5	74,8	78,5	**84,9**	73,7	75,9	75,5

Згідно з результатами, наведеними в табл. 1, використання запропонованого методу виділення адаптивних вейвлет-ознак розпізнавання мовних сигналів на основі оптимізації базової вейвлет-функції за критерієм мінімізації помилки апроксимації дає змогу покращити точність класифікації акустично схожих фонем мови в середньому на 4,1 %, що особливо помітно у випадку таких схожих фонем, як [м], [н] (покращення на 8,3 %). З цього можна зробити висновок, що отримані в результаті оптимізації вейвлет-функції дають змогу побудувати достовірніший набір вейвлет-ознак розпізнавання мови для фонем, які не мають гармонійної структури. Водночас для голосних фонем, що мають гармонійну структуру, запропонований метод підвищує результати

© О. Ю. Лавриненко, 2025

класифікації, в порівнянні з відомими методами, не так суттєво. Цей результат є особливо значущим, оскільки приголосні фонеми становлять семантичний каркас мовного повідомлення. Варто зазначити, що для досягнення порівняльної точності класифікації запропонований метод дає змогу використовувати вейвлет-ознаки розпізнавання мови в 1,5 рази меншої розмірності, ніж МЧКК.

Наприкінці потрібно пояснити основні питання, які впливають на швидкодію паралельного генетичного алгоритму (ПГА) в задачах оптимізації цільової функції мовного сигналу [13].

Розрахунок швидкодії ПГА – це комплексне завдання, яке залежить від безлічі факторів, включаючи архітектуру ПГА, кількість обчислювальних вузлів (ядер процесора) і, в даному випадку, від особливостей цільової функції та вхідних даних (мовного сигналу).

Перш ніж приступити до розрахунків, визначимо основні параметри, такі як, тривалість мовного сигналу (T): 20 мс = 0,02 с; частота дискретизації (Fs): 8 кГц = 8000 Гц.

Першим кроком є визначення кількості відліків (N) у мовному сигналі. Це впливає на розмірність хромосоми в генетичному алгоритмі: $N = T \times Fs = 0.02 \times 8000 = 160$ відліків.

Таким чином, кожен індивідуум у популяції (хромосома) складатиметься з 160 елементів, що представляють собою параметри для оптимізації цільової функції, пов'язаної з цим відрізком сигналу.

Швидкодія (прискорення) паралельного алгоритму (S) зазвичай розраховується за законом Амдала або законом Густафсона. Тут ми використовуємо закон Амдала, оскільки він більш консервативно оцінює потенційне прискорення.

Закон Амдала [14]:

$$S = \frac{1}{(1-p) + \frac{p}{N_{proc}}},$$

де S – прискорення, або відношення часу виконання послідовного алгоритму до часу виконання паралельного; p – частка програми, яка може бути розпаралелена (паралелізована частина); N_{proc} – кількість обчислювальних вузлів (процесорів).

Припустимо, що частка розпаралелюваної частини $p = 0,95$ (95 % обчислень можна розпаралелити), кількість обчислювальних вузлів $N_{proc} = 4$.

Використовуємо закон Амдала для розрахунку прискорення:

$$S = \frac{1}{(1-0,95) + \left(\frac{0,95}{4}\right)} \approx 3,48.$$

Це означає, що теоретичне прискорення в даному випадку складе приблизно 3.48 рази.

Щоб оцінити реальний час, нам потрібно знати час виконання одного покоління для послідовного алгоритму (T_{seq}).

$$T_{seq} = N_{gen} \times \left(N_{pop} \times T_{fitness} + T_{other} \right),$$

де N_{gen} – кількість поколінь; N_{pop} – розмір популяції; $T_{fitness}$ – час обчислення цільової функції для одного індивіда; T_{other} – час на інші операції (селекція, схрещування, мутація).

Якщо ми знаємо T_{seq}, то час виконання паралельного алгоритму (T_{par}) можна оцінити так:

$$T_{par} = \frac{T_{seq}}{S}.$$

Математичний розрахунок швидкодії ПГА для мовного сигналу тривалістю 20 мс і частотою дискретизації 8 кГц показує, що теоретичне прискорення може бути значним, якщо частка розпаралеленої частини алгоритму висока. Конкретне значення швидкодії залежить від кількості обчислювальних вузлів і частки розпаралеленості. У наведеному прикладі, при 4 вузлах і 95% розпаралеленості, прискорення становить 3.48 рази. Однак реальна швидкодія буде також залежати від витрат на комунікацію і синхронізацію.

Час обчислення цільової функції ($T_{fitness}$) для одного індивіда залежить від складності самої функції. Припустимо, що цільова функція – це проста математична операція, така як середньоквадратичне відхилення (СКВ), яка вимагає $O(N)$ операцій, де N – кількість відліків [15].

Припустимо, що час виконання однієї операції (наприклад, додавання, множення) на процесорі становить t_{op}

$$T_{fitness} = N \times t_{op} \times k,$$

де k - коефіцієнт, що відображає складність цільової функції. Наприклад, для СКВ він може дорівнювати 3 (одна операція віднімання, одна операція піднесення до квадрата і одна операція додавання для кожного відліку). Якщо $t_{op} = 1 \times 10^{-9}$ с, то $T_{fitness} = 160 \times 1 \times 10^{-9} \times 3 \approx 0,48$ мкс. Це час обчислення для одного індивіда.

Щоб розрахувати час оптимізації, потрібно визначити час, що витрачається на обчислення цільової функції, яке є найбільш трудомісткою

© О. Ю. Лавриненко, 2025

частиною генетичного алгоритму, а потім застосувати закон Амдала для обліку паралельних обчислень.

Спочатку визначимо час, необхідний для обчислення цільової функції для всієї популяції в одному поколінні.

Кількість відліків (N): 160 (з попередніх розрахунків).

Час обчислення для одного індивідуума ($T_{fitness}$): 0.48 мкс (з попередніх розрахунків).

Розмір популяції (N_{pop}): 25000 індивідуумів.

Час обчислення для всієї популяції в одному поколінні (T_{pop}) на одному ядрі:

$$T_{pop} = N_{pop} \times T_{fitness} =$$
$$= 25000 \times 0.48 \text{ мкс} = 12000 \text{ мкс} = 12 \text{ мс}.$$

Тепер розрахуємо загальний час виконання алгоритму на одному ядрі, враховуючи кількість поколінь ($N_{gen} = 150$):

$$T_{total\_seq} = N_{gen} \times T_{pop} =$$
$$= 150 \times 12 \text{ мс} = 1800 \text{ мс} = 1{,}8 \text{ с}.$$

Це час виконання алгоритму, якби він працював послідовно.

Для розрахунку часу на декількох ядрах ми використовуємо закон Амдала. Припустимо, що частка розпаралеленої частини алгоритму $p = 0{,}95$.

Час виконання паралельного алгоритму (T_{par}) на N_{proc} ядрах:

$$T_{par} = T_{total\_seq} \times \left((1-p) + \frac{p}{N_{proc}} \right).$$

Підставимо значення для кожного випадку:

1 ядро:

$$N_{proc} = 1,$$
$$T_{par\_1} = 1{,}8 \text{ с} \times \left((1-0{,}95) + \frac{0{,}95}{1} \right) = 1{,}8 \text{ с}.$$

8 ядер:

$$N_{proc} = 8,$$
$$T_{par\_8} = 1{,}8 \text{ с} \times \left((1-0{,}95) + \frac{0{,}95}{8} \right) = 0{,}304 \text{ с}.$$

16 ядер:

$$N_{proc} = 16,$$
$$T_{par\_16} = 1{,}8 \text{ с} \times \left((1-0{,}95) + \frac{0{,}95}{16} \right) = 0{,}197 \text{ с}.$$

32 ядра:

$$N_{proc} = 32,$$
$$T_{par\_32} = 1{,}8 \text{ с} \times \left((1-0{,}95) + \frac{0{,}95}{32} \right) = 0{,}143 \text{ с}.$$

64 ядер:

$$N_{proc} = 64,$$
$$T_{par\_64} = 1{,}8 \text{ с} \times \left((1-0{,}95) + \frac{0.95}{64} \right) = 0{,}117 \text{ с}.$$

Розрахунки показують, що використання паралельних обчислень значно скорочує час оптимізації. Збільшення кількості ядер з 1 до 64 скорочує час виконання більш ніж у 15 разів, з 1,8 с до 0,117 с. При цьому, із зростанням кількості ядер, приріст швидкості стає менш помітним через частку непаралельної частини алгоритму (рис. 3).

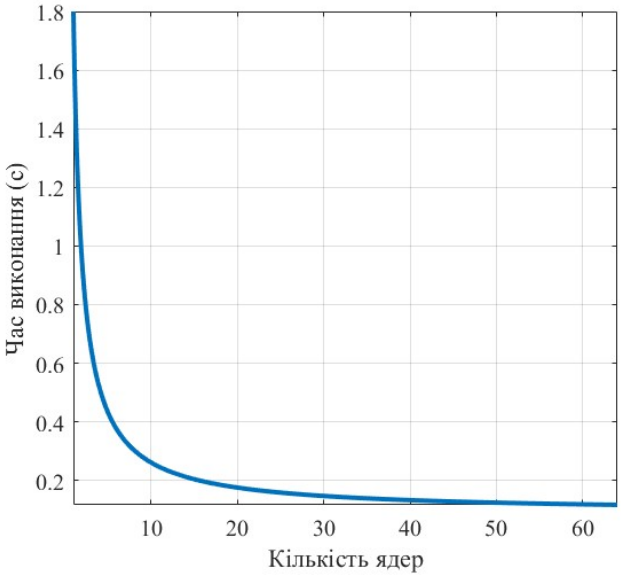

Рис. 3. Залежність часу виконання від кількості ядер

Графік показують, що зі збільшенням кількості ядер процесора час виконання стрімко зменшується спочатку, але потім зменшення сповільнюється, що і є ключовим наслідком закону Амдала.

Висновки

У цій статті запропоновано метод оптимізації базової вейвлет-функції за критерієм мінімізації помилки апроксимації для виділення адаптивних вейвлет-ознак розпізнавання мовних сигналів із використанням генетичного алгоритму та методу опорних векторів. Розроблений метод аналізу складається з таких етапів: 1) побудова адаптивної базової вейвлет-функції для досліджуваного класу мовних сигналів із використанням генетичного алгоритму; 2) проведення вейвлет-аналізу конкретного мовного сигналу, з використанням адаптивної базової вейвлет-функції. Показано, що використання базових вейвлет-функцій, згенерованих на основі представленого методу, для побудови адаптивних вейвлет-ознак розпізнавання мови дає змогу отримати стійке підвищення точності класифікації. Зокрема, в порівнянні з МЧКК поліпшення точності становить від 4,1 до

© О. Ю. Лавриненко, 2025

8,3 % для різних фонем мови. Показано, що згенерована адаптивна вейвлет-функція зберігає інформацію про динамічні процеси в заданому мовному сигналі, і, як наслідок, у 1,5 рази скорочується розмірність вейвлет-ознак розпізнавання мови без втрати точності класифікації.

Таким чином, резюмуючи отримані результати чисельних експериментів, можна зробити висновок, що використання запропонованого методу дає змогу досягти кращих характеристик класифікації фонем мови, ніж використання класичного аналізу на основі МЧКК у 1.4 рази.

ЛІТЕРАТУРА

[1] K. T. Shivaram, N. Mahesh Kumar, M. Anusha and B. S. Manohar, “The optimal numerical wavelet based integration of probability density function by chebyshev wavelet method,” *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, Madurai, India, 2019, pp. 175–177, doi: 10.1109/ICCS45141.2019.9065515.

[2] K. Geetha and M. K. Hota, “Seismic Random Noise Attenuation Using Optimal Empirical Wavelet Transform With a New Wavelet Thresholding Technique,” in *IEEE Sensors Journal*, vol. 24, no. 1, pp. 596–606, 1 Jan.1, 2024, doi: 10.1109/JSEN.2023.3334819.

[3] C. Candemir, “Selection of Wavelet Based Optimal Denoising Method in fMRI Signals,” *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Istanbul, Turkey, 2020, pp. 1–5, doi: 10.1109/ASYU50717.2020.9259816.

[4] O. Lavrynenko, B. Chumachenko, M. Zaliskyi, S. Chumachenko and D. Bakhtiiarov, “Method of Remote Biometric Identification of a Person by Voice based on Wavelet Packet Transform,” *CEUR Workshop Proceedings*, 2024, vol. 3654, pp. 150–162.

[5] G. Yang, Y. Song and J. Du, “Speech Signal Denoising Algorithm and Simulation Based on Wavelet Threshold,” *2022 4th International Conference on Natural Language Processing (ICNLP)*, Xi’an, China, 2022, pp. 304–309, doi: 10.1109/ICNLP55136.2022.00055.

[6] O. Lavrynenko, D. Bakhtiiarov, V. Kurushkin, S. Zavhorodnii, V. Antonov and P. Stanko, “A method for extracting the semantic features of speech signal recognition based on empirical wavelet transform,” *Radioelectronic and Computer Systems*, 2023, vol. 107, no. 3, pp. 101–124. doi: 10.32620/reks.2023.3.09.

[7] Y. Tverdokhlib and V. Dubrovin, “Research on Wavelet Filter Features for Nonstationary Signals,” *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)*, Kyiv, Ukraine, 2019, pp. 785–788, doi: 10.1109/PICST47496.2019.9061501.

[8] O. Veselska, O. Lavrynenko, R. Odarchenko, M. Zaliskyi, D. Bakhtiiarov, M. Karpinski and S. Rajba, “A Wavelet-Based Steganographic Method for Text Hiding in an Audio Signal,” *Sensors*, 2022, vol. 22, no. 15, pp. 1–25. doi: 10.3390/s22155832.

[9] H. Xiao, D. Hu and J. Wang, “Threshold selection of wavelet denoising based on optimization algorithms,” *2022 International Conference on Innovations and Development of Information Technologies and Robotics (IDITR)*, Chengdu, China, 2022, pp. 88–92, doi: 10.1109/IDITR54676.2022.9796485.

[10] S. Sun, Y. Sun, Y. Li, S. Ma, L. Zhang and Y. Hu, “An Adaptive Wavelet Multilevel Soft Threshold Algorithm for Denoising Partial Discharge Signals,” *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, Beijing, China, 2019, pp. 1874–1878, doi: 10.1109/iSPEC48194.2019.8975273.

[11] O. Lavrynenko, R. Odarchenko, G. Konakhovych, A. Taranenko, D. Bakhtiiarov and T. Dyka, “Method of Semantic Coding of Speech Signals based on Empirical Wavelet Transform,” *2021 IEEE 4th International Conference on Advanced Information and Communication Technologies (AICT)*, Lviv, Ukraine, 2021, pp. 18–22, doi: 10.1109/AICT52120.2021.9628985.

[12] O. Yu. Lavrynenko, D. I. Bakhtiiarov, B. S. Chumachenko, O. G. Holubnychyi, G. F. Konakhovych and V. V. Antonov, “Application of Daubechies wavelet analysis in problems of acoustic detection of UAVs,” *CEUR Workshop Proceedings*, 2024, vol. 3662, pp. 125–143.

[13] J. Jiang, Z. Sun, R. Lu, L. Pan and Z. Peng, “Real Relative Encoding Genetic Algorithm for Workflow Scheduling in Heterogeneous Distributed Computing Systems,” in *IEEE Transactions on Parallel and Distributed Systems*, vol. 36, no. 1, pp. 1–14, Jan. 2025, doi: 10.1109/TPDS.2024.3492210.

[14] Y. Yu *et al*., “Memristor Parallel Computing for a Matrix-Friendly Genetic Algorithm,” in *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 5, pp. 901–910, Oct. 2022, doi: 10.1109/TEVC.2022.3144419.

[15] M. Askari, S. Soleimani, M. H. Shakoor and M. Momeni, “Eliminating Network Depth: Genetic Algorithm for Parameter Optimization in CNNs,” in *IEEE Access*, vol. 13, pp. 22473–22488, 2025, doi: 10.1109/ACCESS.2025.3533976.

© О. Ю. Лавриненко, 2025

Лавриненко О. Ю.

МЕТОД ОПТИМІЗАЦІЇ БАЗОВОЇ ВЕЙВЛЕТ-ФУНКЦІЇ ЗА КРИТЕРІЄМ МІНІМІЗАЦІЇ ПОМИЛКИ АПРОКСИМАЦІЇ МОВНИХ СИГНАЛІВ

В роботі представлено розроблений метод виділення адаптивних вейвлет-ознак розпізнавання мовних сигналів на основі оптимізації базової вейвлет-функції за критерієм мінімізації помилки апроксимації. Для побудови базової вейвлет-функції було запропоновано використовувати генетичний алгоритм, у якому в якості цільової функції виступає точність класифікації мовних сигналів методом опорних векторів. Головною перевагою побудованої адаптивної вейвлет-функції за допомогою запропонованого методу є той факт, що вона враховує динаміку мовного сигналу впродовж всього часу, оскільки на кожному масштабі будується вейвлет-функція, яка залежить від усього сигналу та містить у собі інформацію про всі його зміни, що безпосередньо впливає на покращення роздільної здатності вейвлет-ознак розпізнавання мови. Аналіз мовних сигналів запропонованим методом буде складатися із наступних кроків: 1) визначення метакласу мовного сигналу; 2) вибір оптимальної пари «адаптивна вейвлет-функція - класифікатор»; 3) вейвлет-аналіз мовного сигналу з використанням побудованої адаптивної вейвлет-функції; 4) підсумкова класифікація адаптивних вейвлет-ознак розпізнавання мовних сигналів. Показано, що використання базових вейвлет-функцій, згенерованих на основі представленого методу для виділення адаптивних вейвлет-ознак розпізнавання мовних сигналів дає змогу отримати стійке підвищення точності класифікації. Зокрема, в порівнянні з мел-частотними кепстральними коефіцієнтами покращення точності становить від 4.1 до 8.3 % для різних фонем української мови. Показано, що згенерована адаптивна вейвлет-функція має властивість збереження інформації про динамічні процеси в мовному сигналі, і, як наслідок, у 1.5 рази скорочується розмірність адаптивних вейвлет-ознак розпізнавання без втрати точності класифікації. Показано, що використання запропонованого методу дає змогу досягти кращих характеристик класифікації фонем мови, ніж з використанням класичного аналізу на основі мел-частотних кепстральних коефіцієнтів у 1.4 рази.

Ключові слова: мовні сигнали; адаптивні вейвлет-ознаки розпізнавання мови; вейвлет-перетворення; оптимізація вейвлет-функції; мінімізація помилки апроксимації; метод опорних векторів; генетичний алгоритм.

Lavrynenko O.

A METHOD FOR OPTIMIZING THE BASIC WAVELET FUNCTION BY THE CRITERION OF MINIMIZING THE ERROR OF APPROXIMATION OF SPEECH SIGNALS

The paper presents a developed method for selecting adaptive wavelet features for speech signal recognition based on optimizing the base wavelet function by the criterion of minimizing the approximation error. To build the base wavelet function, it was proposed to use a genetic algorithm, in which the accuracy of speech signal classification by the support vector method is the objective function. The main advantage of the constructed adaptive wavelet function using the proposed method is the fact that it takes into account the dynamics of the speech signal throughout time, since a wavelet function is built at each scale that depends on the entire signal and contains information about all its changes, which directly affects the improvement of the resolution of wavelet features of speech recognition. The analysis of speech signals by the proposed method will consist of the following steps: 1) determination of the speech signal metaclass; 2) selection of the optimal pair "adaptive wavelet function - classifier"; 3) wavelet analysis of the speech signal using the constructed adaptive wavelet function; 4) final classification of adaptive wavelet features for speech signal recognition. It is shown that the use of the basic wavelet functions generated on the basis of the presented method for the selection of adaptive wavelet features for speech signal recognition allows to obtain a steady increase in classification accuracy. In particular, in comparison with the mel-frequency capstral coefficients, the accuracy improvement ranges from 4.1 to 8.3 % for different phonemes of the Ukrainian language. It has been shown that the convolved adaptive wavelet function has the property of preserving information about dynamic processes in the speech signal, and, as a result, the dimensionality of adaptive wavelet recognition features is reduced by 1.5 times without loss of classification accuracy. It is shown that the use of the proposed method allows achieving better speech phoneme classification characteristics than using the classical analysis based on mel-frequency cepstral coefficients by a factor of 1.4.

Keywords: speech signals; adaptive wavelet features for speech recognition; wavelet transform; wavelet function optimization; minimization of approximation error; support vector method; genetic algorithm.

Стаття надійшла до редакції 03.06.2025 р.
Прийнято до друку 03.09.2025 р.

© О. Ю. Лавриненко, 2025