

УДК 621.316.11

А. В. Волошко, Т. М. Лутчин, Д. Ю. Терещенко

НТУУ «КПІ» ІЕЕ, м. Київ, Україна

## МЕТОД КЛАСИФІКАЦІЇ ІНФОРМАЦІЙНИХ СИГНАЛІВ ЗА ДОПОМОГОЮ ПОБУДОВИ ОПТИМАЛЬНОГО ДЕРЕВА ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ

*Пропонується побудувати оптимальне дерево вейвлет-розкладу на основі аналізу функції вартості – ентропії вузлів. Це підвищує швидкодію алгоритму віднесення ГЕН до певного класу. Для побудови збалансованого дерева вейвлет-перетворення застосовуються пакетні вейвлети.*

*Предлагается построить оптимальное дерево вейвлет-разложения на основе анализа функции стоимости – энтропии узлов. Это увеличивает скорость алгоритма отнесения ГЕН к определенному классу. Для построения сбалансированного дерева вейвлет-преобразования используются пакетные вейвелеты.*

*It is proposed to construct an optimal tree wavelet decomposition based on an analysis of the cost function – the entropy of nodes. This increases the speed of the algorithm for inclusion of graphs of electrical loads to a certain class. To construct a balanced tree of wavelet packet transform uses wavelets.*

Як відомо, згідно з теорією вейвлет-перетворення будь-яку послідовність дискретних вибірок  $x(t_i)$  можливо представити у вигляді упорядкованої сукупності їх коефіцієнтів розкладання по системі масштабних функцій  $\varphi_{m,k}$  та вейвлетів  $\psi_{m,k}$ . Масштабні функції та вейвлети визначаються відповідно з теорією КМА. Маллатом [1] запропоновано для розрахунку апроксимуючих  $a_{m,k}$  та деталізуючих коефіцієнтів  $d_{m,k}$  застосовувати схему послідовного поділу, яка інтерпретується як послідовна двуполосна фільтрація вхідного сигналу за допомогою каскадно-з'єднаних блоків фільтрів низької та високої частоти.

При цьому каскадування проводиться по низькочастотній області. Причина цього – в припущенні, що ця область містить більше інформації про сигнал. В результаті отримуємо так зване “однобоке дерево”. На практиці, для багатьох реальних сигналів дане твердження не зовсім вірне, і тому застосовуються так звані пакетні вейвлети [2]. При їх використанні операції послідовного частотного перетворення застосовуються як для низькочастотних, так і для високочастотних коефіцієнтів. В результаті отримуємо збалансоване дерево вейвлет-перетворення – бінарне дерево.

При застосуванні пакетних вейвлетів вхідний сигнал описується трьома параметрами: позиції та масштабу (як для звичайного вейвлет-перетворення), а також частотою.

Як відомо, таке пакетне вейвлет-перетворення являється адаптивним. Ця адаптація не потребує навчання або відомостей про статистичні властивості сигналів і дозволяє більш точно враховувати особливості аналізованого сигналу шляхом вибору відповідної оптимальної форми дерева розкладу. В процесі дослідження це дозволило забезпечити мінімальну кількість вейвлет-коефіцієнтів при заданій точності стискання/відновлення сигналів, їх класифікації.

Дослідження проводилось в плані розробки метода класифікації та стискання/відновлення ГЕН за допомогою побудови оптимального дерева тому, що, по-перше, бінарні дерева доволі просто можуть бути представлені у вигляді списків, або масивів. При цьому кожний елемент дерева має поле даних і два поля покажчиків. Один покажчик зв'язує елемент з правим нащадком, а другий – з лівим.

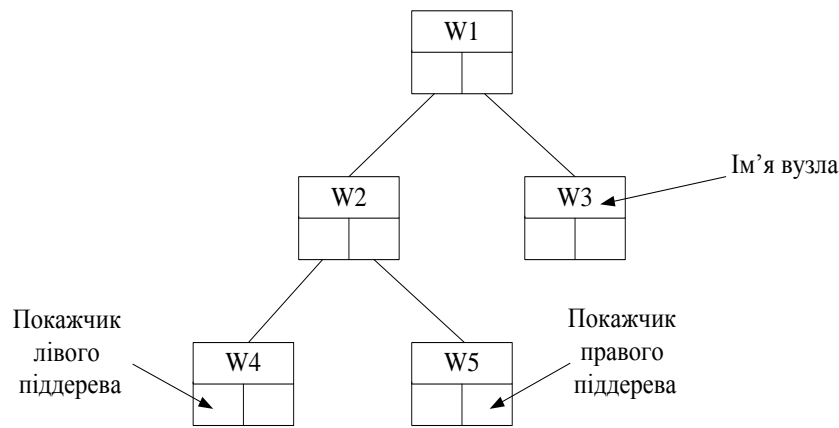


Рисунок 1. Представлення бінарного дерева у вигляді спискової структури.

Дослідження показали, що представлення бінарного дерева у вигляді масиву являється більш ефективним в нашому випадку. А саме, бінарне дерево завжди має строго означене число вершин на кожному рівні, які нумеруються зліва направо послідовно по рівням і використовуються в якості індексів в одновірних масивах.

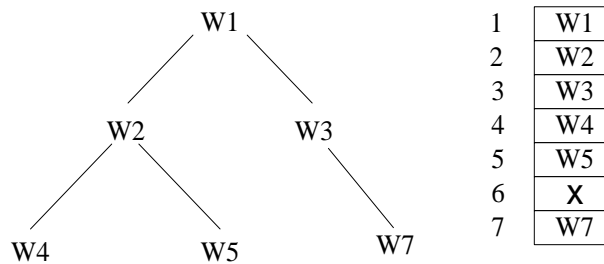


Рисунок 2. Представлення бінарного дерева у вигляді масиву.

Також, в зв'язку з тим, що в процесі обробки число рівнів дерева суттєво не змінюється, то такий спосіб представлення бінарного дерева значно більш економічний на відміну від стекової структури. При цьому, адреса будь-якої вершини в одновірному масиві обчислюється, як

$$agp = (2k - 1) + (i - 1),$$

де  $k$  – номер рівня вершини,  $i$  – номер на рівні  $k$  в повному бінарному дереві.

Дослідження показали, що головним недоліком даного способу представлення бінарного дерева являється те, що структура даних є статичною. Розмір масиву вибирається виходячи із максимально можливої кількості рівнів бінарного дерева. Виходячи з цього чим менш повним являється дерево, тим менш раціонально використовується пам'ять.

Як було відмічено раніше, за допомогою побудови оптимального дерева декомпозиції при проведенні зворотної вейвлет-декомпозиції (реконструкція сигналу) незначні, інформаційно-надлишкові, або зовсім непотрібні деталі сигналів виключаються. Мірою оптимальності при цьому являється кількість апроксимуючих і деталізуючи коефіцієнтів для реконструкції сигналу при заданій похибці.

В якості функції вартості інформативності набору вейвлет-коефіцієнтів використовується ентропія Шенона. Дана функція буде великою, якщо вейвлет-коефіцієнти приблизно однієї величини, і невеликою, якщо всі вейвлет-коефіцієнти, окрім декількох, близькі до нуля. Під ентропією розуміється величина

$$E = \exp \left( - \sum_{K=1}^N P_K \log_2 \left( \frac{P_K}{2} \right) \right),$$

де  $x$  – вхідний сигнал, вейвлет-коефіцієнти  $\left\{ x_k \right\}_{1 \leq k \leq N}$  якого зустрічаються з ймовірністю  $\left\{ p_k \right\}_{1 \leq k \leq N}$ .

Побудова оптимального дерева декомпозиції проводилась шляхом визначення ентропії вузлів і його апроксимуючих і деталізуючих коефіцієнтів – нащадків. Якщо ентропія вузла була менше ентропії нащадків, подальша декомпозиція в даному вузлі закінчувалась і дерево “обрізалось”. Алгоритм декомпозиції рекурсивно продовжувався до досягнення глибини декомпозиції.

Практична реалізація:

Для побудови вектора ознак класифікації ГЕН на основі оптимального дерева декомпозиції були проаналізовані добові ГЕН за період лютий – червень 2011 р. Було виділено три класи: вихідні, святкові дні та робочі дні. В межах одного класу структури оптимальних дерев декомпозиції були однакові. Порівняння проводились за допомогою послідовних алгоритмів та алгоритму Вітербі [3, 4]. Види оптимальних дерев відповідних класів представлені на рисунку 3.

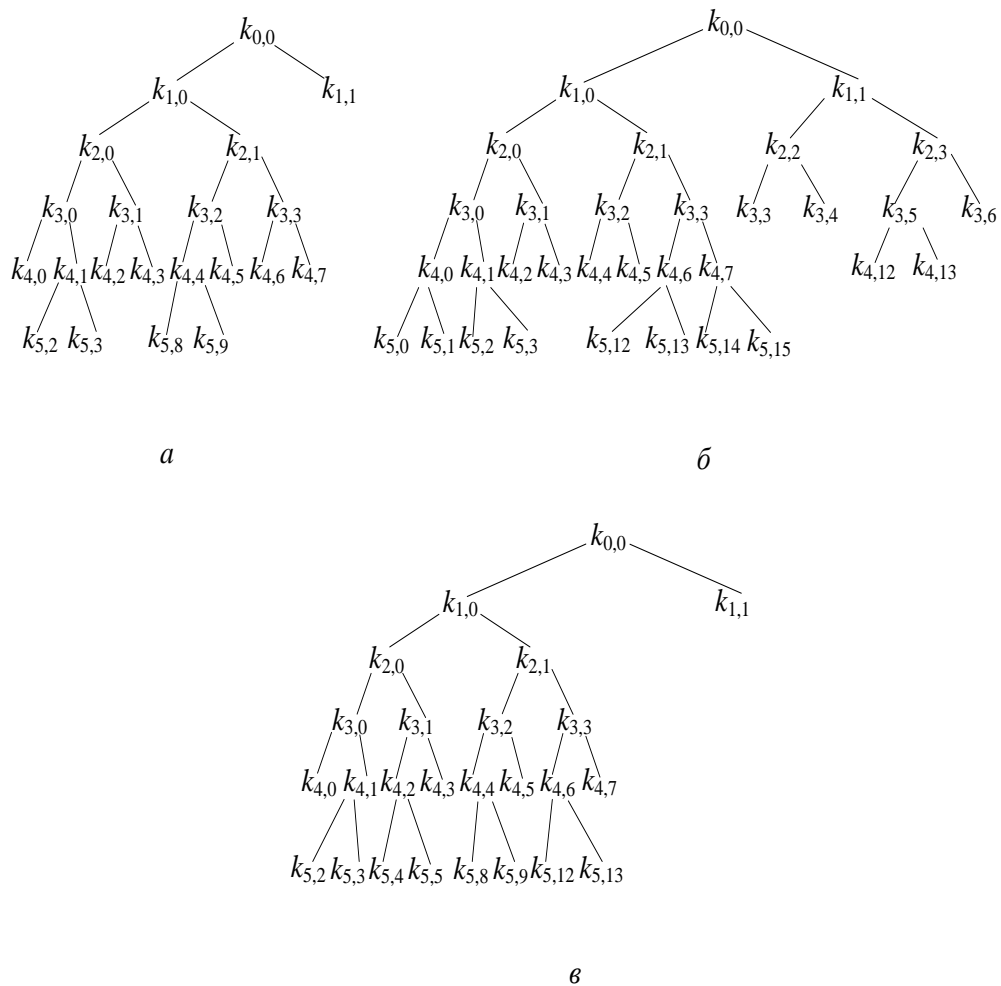


Рисунок 3. Оптимальні дерева: а) – вихідні, б) – святкові, в) – робочі

Більш ефективним алгоритмом віднесення ГЕН до відповідного класу виявився стековий алгоритм. Алгоритм віднесення ГЕН до одного із трьох класів представлений на рисунку 4.

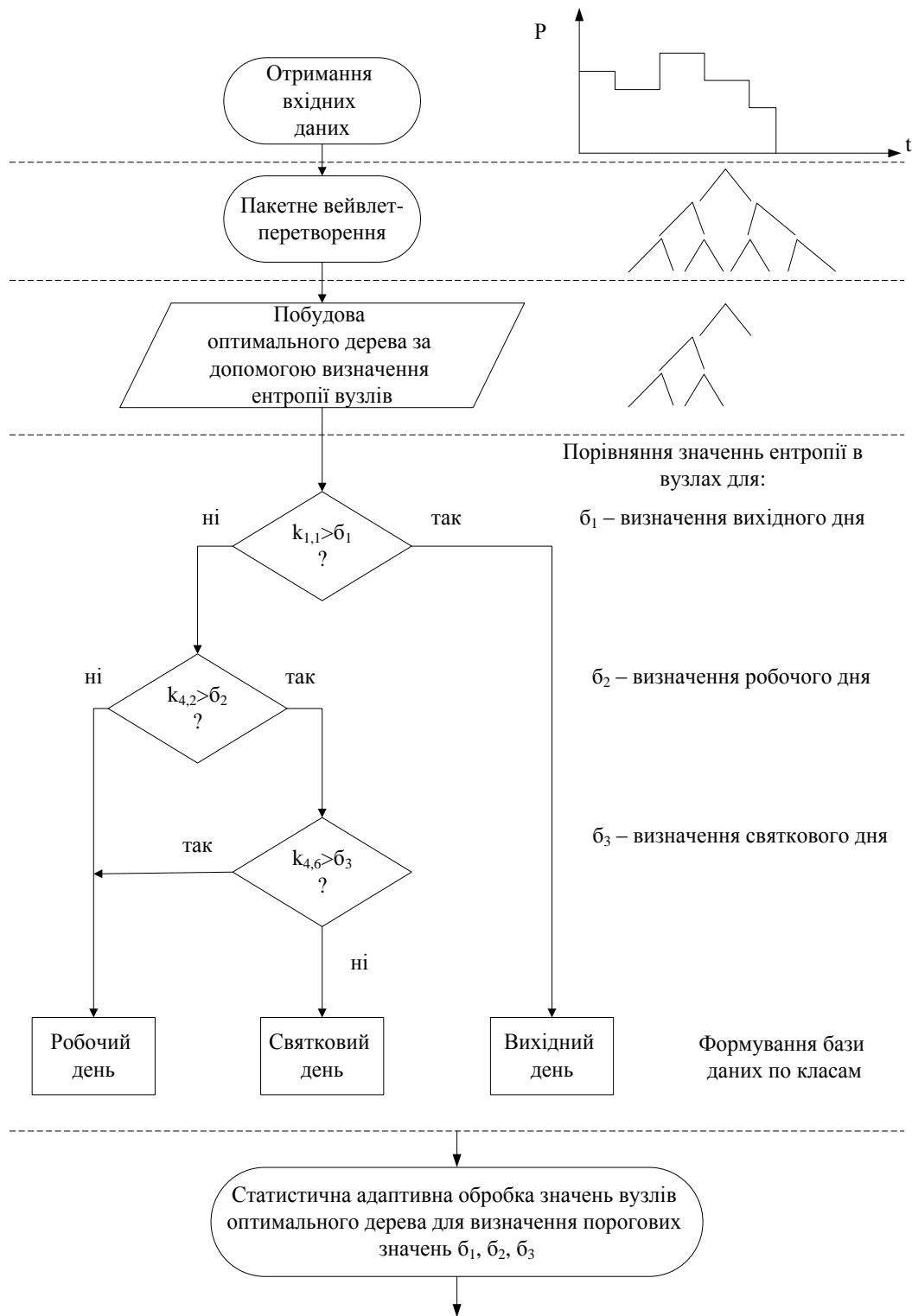


Рисунок 4.

## Алгоритм віднесення ГЕН до одного з класів

Порогові значення  $\delta_1, \delta_2, \delta_3$  – адаптуються до технологічного процесу, характеру виробництва на стадії попередньої статистичної обробки реальних графіків.

Висновки:

В результаті проведених досліджень необхідно відмітити наступне. Пакетне вейвлет-перетворення являється адаптивним і дозволяє більш точно виявляти особливості сигналів шляхом вибору відповідної оптимальної форми дерева декомпозиції. Дана оптимальна форма забезпечує мінімальну кількість

вейвлет-коефіцієнтів при заданій точності реконструкції сигналу. Класифікація ГЕН проводилась за допомогою послідовних алгоритмів та алгоритму Вітербі. Як найбільш оптимальний обраний стековий алгоритм. Для підвищення швидкодії алгоритму класифікації на основі аналізу функції вартості – ентропії вузлів і розділених гілок дерева побудовано оптимальне дерево вейвлет-розкладу, яке дозволило також зменшити об'єм пам'яті, необхідний для зберігання даних.

- [1] **Mallat S. G.** Trans. Am. Math. Soc. 315 69 (1989); IEEE Trans. Patt. Anal. Mach. Int. 31 674 (1989)
- [2] **Strang G., Nguyen T.** Wavelets and Filters Banks. - Wellesley-Cambridge-Press 1996. - 490 p.
- [3] **Витерби А. Д., Омура Дж. К.** Принципы цифровой связи и кодирования. М.: Радио и связь, 1982.
- [4] **Зигангиров К. Ш.** Процедуры последовательного декодирования. М.: Связь, 1974.