

УДК 004.89:[519.812-004.89+519.816-004.85]

ПОРІВНЯННЯ МЕТОДІВ ВИЯВЛЕННЯ ТРЕНДУ НА ЧАСОВИХ РЯДАХ МАЛОЇ ДОВЖИНИ

Яндуков Д.Є.

Науковий керівник – д-р техн. наук, проф. Кіріченко Л.О.
Харківський національний університет радіоелектроніки, каф. ПМ
м. Харків, Україна

тел. (057) 702-14-36, email: dmytro.yandukov@nure.ua

This work is devoted to the assessment of the criteria for determining the presence of a trend in short time series, developed on the basis of the k -means method and the method of determining the trend component using the discrete wavelet transformation, and comparing the performance of these methods. A software product was developed to test the hypothesis of no trend and compile static data on the performance of both methods. An analysis was performed to compare the performance of both methods to identify key differences between the methods.

Одним з найбільш важливих завдань при дослідженні та аналізі коротких часових рядів є виявлення та статистична оцінка основної тенденції розвитку досліджуваного процесу. Головним завданням проведеного дослідження є детектування тренду за найменшою кількістю значень часового ряду і на основі цього проведення кластеризації часових рядів методами машинного навчання.

Запишемо множину часових рядів, які потрібно кластеризувати за наявністю трендової складової, у вигляді:

$$S = \{(S_1^0, \dots, S_{N^0}^0), (S_1^1, \dots, S_{N^1}^1), \dots, (S_1^i, \dots, S_{N^i}^i), \dots, (S_1^n, \dots, S_{N^n}^n)\}, \quad (1)$$

де $(S_1^i, \dots, S_{N^i}^i)$ – часові ряди малої довжини, $i = 1, n$; n – кількість об'єктів, що містяться у файлі з часовими рядами; N^i – довжина часового ряду.

Кожен часовий ряд складається з двох складових, а саме, тренда і шуму:

$$S^i = T(\tau) + \xi / SNR, \quad (2)$$

де $\xi(t) \sim N(0, \sigma)$ – білий шум; $T(\tau)$ – тренд, модель якого вибирається з деякої множини $\tau = \{\text{лінійний, поліноміальний, експоненційний, гіперболічний, логарифмічний}\}$; SNR (signal-to-noise ratio) – це безрозмірна величина, яка дорівнює відношенню сигналу і шуму $SNR = \sigma_{T(\tau)} / \sigma_\xi$.

Метод k -середніх є одним із широко використовуваних методів кластеризації. Спочатку ми задаємо кількість кластерів та відповідні центроїди (центри кластерів) для кожного з них. Далі розподіляємо всі об'єкти за кластерами відповідно до близькості до центрів і знаходимо нові центроїди. Продовжуємо процес доти, доки координати центроїдів не перестануть змінюватися.

Дискретне вейвлет-перетворення – реалізація вейвлет-перетворення з

використанням дискретного набору масштабів та переносів вейвлета, які підпорядковуються певним правилам. Набір вейвлетів формує ортогональний базис, який ми використовуємо для розкладання часового ряду на складові.

За допомогою програмного продукту було проведено аналіз для порівняння кластеризації за різними підходами, різними трендами та відношенням SNR . На рис. 1 показано декомпозицію часового ряду (а) і результати кластеризації методом k -середніх (б). Показано, що якщо не потрібно відокремлювати нетипові часові ряди, метод k -середніх показує досить добрі результати; проте метод вейвлет-розкладання більш гнучкий та показує хороший результат під час роботи в різних умовах.

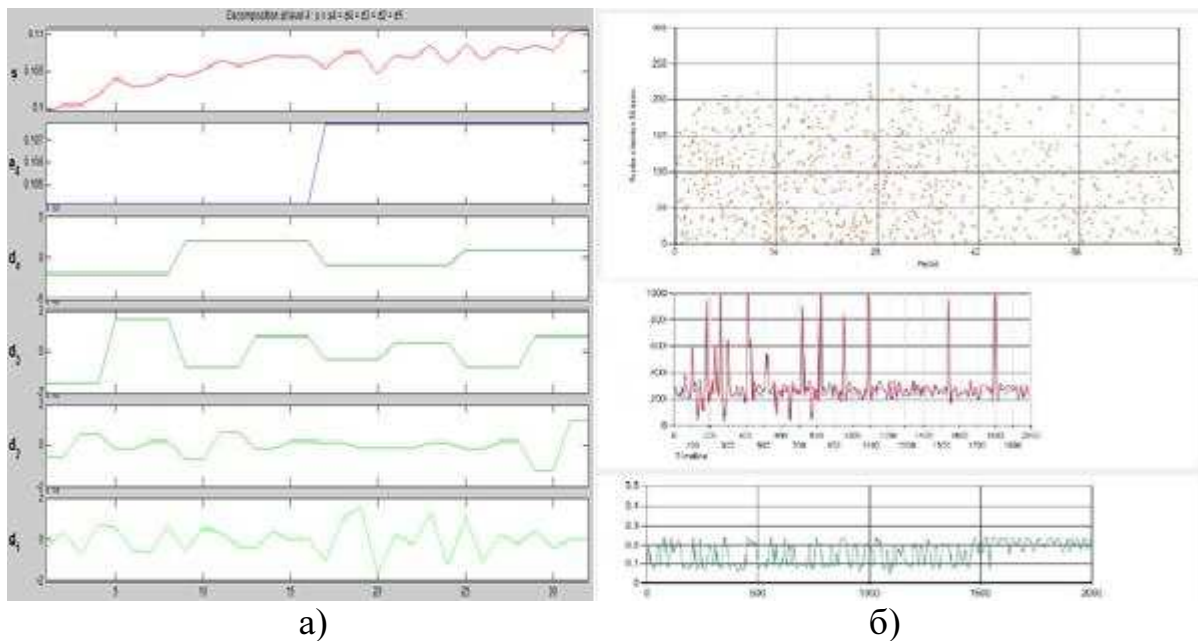


Рисунок 1 – Реалізацій методу: вейвлет-розкладання (а), k -середніх (б)

Список використаних джерел:

1. Кіріченко, Л.О., Ткаченко, А.Є., & Радівілова Т.А. (2019). Кластеризація зашумлених часових рядів. *Системні технології*, 3 (122), 133-139.
2. Зінкевич, І.Е., Кириченко, Л.О., & Радівілова, Т.А. (2017). Порівняльний аналіз методів прогнозування слабкорельованих часових рядів. *Системні технології*, 2 (109), 43-50.