Класифікація Фрактальних Часових Рядів Методами Машинного Навчання

Людмила Кіріченко
кафедра прикладної математики
Харківський національний університет радіоелектроніки
Харків, Україна
lyudmyla.kirichenko@nure.ua

Юлія Кобицька
кафедра прикладної математики
Харківський національний університет радіоелектроніки
Харків, Україна
yuliia.kobytska@nure.ua

Тамара Радівілова
кафедра інфокомунікаційної інженерії
Харківський національний університет радіоелектроніки
Харків, Україна
tamara.radivilova@nure.ua

Classification of Fractal Time Series by Machine Learnin Methods

Lyudmyla Kirichenko
Department of Applied Mathematics
Kharkiv National University of Radioelectronics
Kharkiv, Ukraine
lyudmyla.kirichenko@nure.ua

Yuliia Kobytska
Department of Applied Mathematics
Kharkiv National University of Radioelectronics
Kharkiv, Ukraine
yuliia.kobytska@nure.ua

Tamara Radivilova
Department of Infocommunication Engineering
Kharkiv National University of Radioelectronics
Kharkiv, Ukraine
tamara.radivilova@nure.ua

***Анотація*—У статті проведено порівняльний аналіз класифікації фрактальних часових рядів за допомогою мета-алгоритму на основі дерев рішень Random forest. Для побудови модельних фрактальних часових рядів були обрані біноміальні стохастичні каскадні процеси. Результати свідчать про велику перевагу методів машинного навчання перед традиційними методами оцінювання фрактальних характеристик при класифікації часових рядів за фрактальними властивостями.**

***Abstract*—The article compares the classification of fractal time series using a meta-algorithm based on the Random forest decision trees. Binomial stochastic cascade processes were chosen to construct model fractal time series. The results show a great advantage of machine learning methods over traditional methods of estimating fractal characteristics when classifying time series by fractal properties.**

Ключові слова—мультифрактальні часові ряди, класифікація часових рядів, Random forest, показник Херсту.

Keywords—multifractal time series, time series classification, Random forest, Hurst parameter.

I. Вступ

Часові ряди є основною інформацією для розуміння динаміки в реальних складних системах різного типу. Протягом останніх двох десятиліть були запропоновані і розроблені методи інтелектуального аналізу даних часових рядів, в тому числі методи машинного навчання [1-3]. Машинне навчання використовуються для різних завдань аналізу часових рядів, включаючи класифікацію. Для розробки і тестування нових методів необхідно мати бази даних реальних часових рядів, проте, не менш важливі модельні дані з необхідними властивостями [4, 5].

Багато складних систем мають фрактальну структуру, а їх динаміка представлена часовими рядами, які мають фрактальні (самоподібні) властивості. Аналіз фрактальних властивостей часових рядів широко використовується в різних областях знань [6]. У багатьох випадках виникають проблеми розпізнавання і класифікації фрактальних часових рядів. Найчастіше такі завдання вирішуються шляхом оцінки та аналізу фрактальних характеристик в часовому ряду [7-9]. Однак в останні роки зростає інтерес до методів машинного навчання для аналізу і класифікації фрактальних рядів [10-12]. Метою даної роботи є порівняльний аналіз класифікації фрактальних стохастичних часових рядів, які виконуються мета-алгоритмами з використанням методів дерева рішень.

II. Мультифрактальні часові ряди

Стохастичний процес *X(t)* є самоподібним з параметром самоподібності *H*, якщо процес $a^{-H}X(at)$ описується тими ж кінцевомірними законами розподілів, що і *X(t)*:

$$Law\{X(t)\} = Law\{a^{-H}X(at)\}, \vee\ a > 0,\ t > 0\ .$$

Параметр *H*, 0<H<1, називається показником Херста, який являє собою ступінь самоподібності. Початкові моменти самоподібного випадкового процесу мають властивість

$$M\left[|X(t)|^q\right] \propto \cdot t^{qH}\ .$$

Мультифрактальні процеси проявляють більш гнучкі скейлінгові закономірності для моментних характеристик

$$M\left[|X(t)|^q\right] \propto \cdot t^{qh(q)}\ ,$$

де *h(q)* - узагальнений показник Херста, який є нелінійною функцією, для якої значення $h(q=2)=H$ [13].

Простими моделями мультифрактального процесу з заданими властивостями є біноміальні мультиплікативні каскади [13]. При їх побудові початковий одиничний інтервал ділиться на два рівних інтервали, яким приписуються вагові коефіцієнти w_1 і $1\text{-}w_1$, де w_1 є значенням деякої заданої випадкової величини *W*.

На другому кроці додаються два нових незалежних випадкових значення w_2 і w_3. Вийде 4 інтервали з ваговими коефіцієнтами

$$w_1w_2,\ w_1(1-w_2),\ (1-w_1)w_3\ \text{ та }\ (1-w_1)(1-w_3).$$

При $n \to \infty$ ми приходимо до граничної міри, що є неоднорідною фрактальною множиною.

III. Метод мультифрактального детрендованого флуктуаційного аналізу

При дослідженні властивостей мультифрактальних процесів застосовується мультифрактальний флуктуаційний аналіз (МФДФА) [14]. В цьому випадку кумулятивний ряд $y(t)$, розбивається на *N* сегментів довжиною *s*. Для кожного сегменту *y(t)* обчислюється флуктуаційна функція

$$F^2(s) = \frac{1}{s}\sum_{i=1}^{s}\left(y(t) - Y_m(t)\right)^2\ ,$$

де $Y_m(t)$ - *m*-поліноміальний тренд в межах даного сегменту. При проведенні МФДФА досліджується залежність флуктуаційної функції

$$F_q(s) = \left\{\frac{1}{N}\sum_{i=1}^{N}\left[F^2(s)\right]^{\frac{q}{2}}\right\}^{\frac{1}{q}}\ ,$$

отриманої усередненням на усіх сегментах. Якщо досліджуваний ряд має мультифрактальні властивості, то функція $F_q(s)$ представляється ступеневою залежністю

$$F_q(s) \propto s^{h(q)}\ .$$

Оцінка показника Херста може бути представлена інтервалом значень, всередині якого із заданою ймовірністю знаходиться істинне значення

$$H\!:\ \hat{H} + \Delta - t_\alpha S < H < \hat{H} + \Delta + t_\alpha S\ ,$$

де $\hat{H}$ - оцінка показника Херста у реалізації довжини *N*; Δ - величина систематичного зсуву оцінки; S - середньоквадратичне відхилення оцінки; t_α - квантиль нормального розподілу для рівня значимості α [15].

IV. Класифікація методом Random Forest

Одним з найбільш простих і ефективних методів для вирішення задач класифікації, що виникають в різних областях, вважається метод дерев рішень. Він полягає в тому, щоб здійснювати процес розщеплення вихідних даних на групи, поки не будуть отримані однорідні їх підмножини. Сукупність правил, які дають таке розбиття, дозволяє потім робити висновок для нових даних.

Дерева рішень є нестійкими моделями: невелика зміна в навчальній вибірці може привезти до істотних змін в структурі дерева. У цьому випадку доцільно використовувати ансамблі моделей. Одним з найвідоміших видів ансамблів є метод Random Forest [16]. Random Forest - технологія класифікації, де усі елементарні класифікатори є вирішальними деревами, які навчаються і працюють незалежно один від одного. Основна ідея полягає в тому, що вони не виправляють помилки один одного, а компенсують голосуванням або усередненням.

V. Опис експерименту та результати

Класифікація проводилася для мультифрактальних часових рядів, які отримано генерацією стохастичних біноміальних каскадів на основі симетричного бета-розподілу. В цьому випадку, побудова каскаду дозволяє отримати мультифрактальні часові ряди з показником Херста в діапазоні $H \in (0{,}5;1)$ і ряди можна розбити на класи за значеннями показника Херста.

Кожен клас представляє собою набір згенерованих часових рядів з показником Херста, який лежить в заданому діапазоні значень. Для генерації каскадів всередині кожного класу значення *H* вибиралися за допомогою рівномірного закону розподілу. При формуванні класів значення діапазонів показника Херста змінювалися в інтервалі від 0.5 до 1 з кроком 0.05. Таким чином навчання моделей проходило на 11 класах. Для проведення класифікації був обраний метод Random Forest з ансамблями регресійних дерев.

У роботі для визначення приналежності часового ряду до одного з 11 класів були використані три різні підходи. У першому підході класифікація проводилася безпосередньо за значеннями часового ряду методом Random Forest, де об'єктом був каскадний часової ряд, а ознаками - значення цього ряду. При використанні регресійних дерев рішень результатом роботи моделі була ймовірність відповідності мультифрактального каскаду заданому класу.

Другий підхід до класифікації мультифрактальних каскадів також базувався на машинному навчанні, але в цьому випадку для класифікації були використані статистичні та мультифрактальні характеристики, отримані за значеннями часового ряду. В цьому випадку об'єктами були часові каскадні ряди, а ознаками - оцінки характеристик, розраховані для кожного каскаду.

Третій спосіб класифікації заснований на прямому оцінюванні показника Херста за часовим рядом та визначенні довірчого інтервалу. Вибір класу визначався як ймовірність знаходження істинного значення показника Херста в рамках довірчого інтервалу за співвідношенням ширини класу.

В роботі для побудови моделей дерев рішень використовувалася мова Python з бібліотеками, що реалізують методи машинного навчання. Навчання моделей для кожного класу проводилося на 500-х прикладах часових рядів навчання і перевірялося на 50 тестових. У разі застосування довірчих інтервалів класифікація також проводилася для тієї ж тестової вибірки. Класифікація проводилася для часових рядів з різною довжиною, проте для порівняння результатів основна увага була приділена рядам розміром 512 і 4096 значень.

В таблиці представлено середні ймовірності визначення класу в залежності від довжини мультифрактальних рядів і методу класифікації.

ТАБЛИЦЯ I. СЕРЕДНЯ ЙМОВІРНІСТЬ ВИЗНАЧЕННЯ КЛАСУ

Метод класифікації	**Ймовірність**		**Час навчання**	
	512	4096	512	4096
За значеннями часового ряду	0,74	0,82	15 хвилин	120 хвилин
За статистичними і мультифрактальними характеристиками	0,72	0,77	1,2 хвилини	1,8 хвилин
Оцінювання показника Херста	0,18	0,25		

Таким чином, отримані результати свідчать про велику перевагу методів машинного навчання перед традиційними методами оцінювання фрактальних характеристик при класифікації часових рядів за фрактальними властивостями. Однак для навчання необхідна достатня кількість часових рядів із заздалегідь відомими властивостями, що в багатьох задачах є проблематичним. Класифікація, яка проведена на основі навчання за значеннями часового ряду показала трохи більш високу точність, ніж на основі навчання за фрактальними і статистичними характеристиками. Однак необхідно відзначити, що в цьому випадку час навчання моделі нелінійно зростає при зростанні довжини часового ряду.

ЛІТЕРАТУРА REFERENCES

[1] Esling P., Agon C. «Time series data mining», in *ACM Computing Surveys* 46(1), (2012).

[2] Ben D. «Feature-based time-series analysis» [Online]. Aveilable: https://arxiv.org/abs/1709.08055, last accessed 2018/01/26.

[3] Bagnall A., Lines J., Bostrom A., Large J., Keogh E. "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances", in *Data Mining and Knowledge Discovery* 31 (3), 606-660, 2017.

[4] Garima R. "Review on time series databases and recent research trends in Time Series Mining", in *Proceedings of the 5th confluence the next generation information technology summit*, IEEE, 109–115. Noida, India, 2014.

[5] Bagnall A., Bostrom A., Large J., Lines J. "Simulated Data Experiments for Time Series Classification Part 1: Accuracy Comparison with Default Settings" [Onine]. Aveilable: https://arxiv.org/abs/1703.09480v1, last accessed 2018/01/26.

[6] Brambila F. "Fractal Analysis - Applications in Physics, Engineering and Technology" [Online]. Available: https://www.intechopen.com/books/fractal-analysis-applications-in-physics-engineering-and-technology, last accessed 2018/01/26.

[7] Banerjee A., Sanyal S., Guhathakurata T., Sengupta R., Ghosh D. "Categorization of stringed instruments with multifractal detrended fluctuation analysis" [Online]. Available: https://arxiv.org/abs/1601.07709, last accessed 2018/01/26.

[8] Hippenstiel R., El-Kishky H., Radev P. "On time-series analysis and signal classification - part I: fractal dimensions", *38th Asilomar Conference on Signals Systems and Computers,* 2121-2125, 2004.

[9] Korus L., Piórek M. "Compound method of time series classification", in *Nonlinear Analysis: Modelling and Control*, 20(4), 545–560, 2015.

[10] Andre L., Coelho V., Clodoaldo A., Lima M. "Assessing fractal dimension methods as feature extractors for EMG signal classification", in *Engineering Applications of Artificial Intelligence* 36, 81–98, 2014.

[11] Symeonidis S. "Sentiment analysis via fractal dimension", in *Proceedings of the 6th Symposium on Future Directions in Information Access*, 48-50, 2015.

[12] S. P. Arjunan, D. K. Kumar, G. R. Naik "A machine learning based method for classification of fractal features of forearm sEMG using Twin Support vector machines", in *Annual International Conference of the IEEE Engineering in Medicine and Biology*, 4821-4824, 2010.

[13] R.H.Riedi. Multifractal processes, in Doukhan P., Oppenheim G., Taqqu M.S. (Eds.), Long Range Dependence "Theory and Applications", in *Birkhuser*, 625–715, 2002.

[14] J. Kantelhardt, S. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde and H. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series", in *Physica A: Statistical Mechanics and its Applications*, vol. 316, no. 1-4, pp. 87-114, 2002

[15] L. Kirichenko, T. Radivilova and V. Bulakh "Generalized approach to Hurst exponent estimating by time series", in *Informatics Control Measurement in Economy and Environment Protection*, vol. 8, no. 1, pp. 28-31, 2018.

[16] Breiman L. "Random Forests", in *Machine Learning*, 45 (1), 5–32, 2001.