ПРИМЕНЕНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ ХААРА К АНАЛИЗУ СИГНАЛОВ

В.И. Лобач
Кафедра математического моделирования и анализа данных, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: lobach@bsu.by

На осонове вейвлетов Хаара и применения критерия хи-квадрат решается задача обнаружения неоднородности дискретного временного ряда.

Сигнал x_t содержит полную информацию во временном пространстве, т. е. амплитуду сигнала в каждый фиксированный момент времени t. Однако из x_t невозможно напрямую получить информацию, описывающую частотную характеристику сигнала. С другой стороны, разложение $X(\omega)$, полученное с помощью преобразования Фурье сигнала x_t, содержит в себе всю информацию о сигнале в частотном пространстве, т. е. амплитуду и фазовый угол каждой частотной компоненты ω. Но теперь нет никакой информации о том, где конкретно во времени та или иная компонента появилась. Иными словами, ни x_t, ни $X(\omega)$ не дают полного описания сигнала.

Если характеристики сигнала меняются с течением времени незначительно, т. е. сигнал стационарен, тогда преобразование Фурье является удобным инструментом анализа сигнала. Однако во многих приложениях сигнал быстро меняющийся или нестационарный (например, имеется тренд или резкие скачки). В этом случае анализ сигнала с помощью преобразования Фурье не способен определить, где или когда эти изменения имели место. С целью преодоления этих недостатков метода Фурье используется другое преобразование под названием вейвлет-преобразование [1, 2]. Вейвлет-анализ широко применяется к решению различных задач обработки сигналов: обнаружение скачков и нерегулярностей в сигнале, сжатие сигнала, анализ сигналов датчика в робототехнике, стеганографический анализ сигнала и т. д.

В данной работе рассматривается проблема обнаружения неоднородностей двоичного дискретного временного ряда.

Пусть множество $\mathrm{V}=\{0,\,1\}$, $\mathrm{V}^N=\{0,\,1\}^N$ – последовательность длины N, состоящая из 1, 0; $a=\{a_t\in\{0,\,1\},\,t=0,\,N-1\}$ – периодическая с периодом T, $T\ll N$, последовательность детерминированных величин. Наблюдается последовательность вида

$$x=\{x_t|x_t=\xi_t\oplus a_t,\,t=\overline{0,\,N-1}\},$$

где $\xi_t\oplus a_t=\xi_t+a_t(\bmod 2)$.

Требуется по наблюдениям $x=\{x_t,\,t=\,=\,0,\,\overline{N-1}\}$ определить, является ли последовательность $\{x_t\}$, $t\in\overline{0,\,N-1}$, периодической, т. е. ставится задача обнаружения скрытой периодичности.

Математически данную проблему можно сформулировать как задачу проверки гипотез.

Основная гипотеза и альтернативная гипотеза:

H_0: $\{x_t,\,t=\overline{0,\,N-1}\}$ – последовательность независимых одинаково распределенных случайных величин $Bi(1,\,p)$,

H_1: $\{x_t,\,t=\overline{0,\,N-1}\}$ – не является последовательностью независимых одинаково распределенных случайных величин $Bi(1,\,p)$.

Вместо анализа наблюдаемой последовательности $\{x_t,\,t=\overline{0,\,N-1}\}$, будем исследовать статистики, являющиеся коэффициентами дискретного вейвлет-преобразования Хаара исходной последовательности.

При выполнении нулевой гипотезы H_0 имеются данные $X=(x_0,\,x_1,\,\ldots,\,x_{N-1})$, которые представляют собой последовательность независимых одинаковых распределенных случайных величин $Bi(1,\,p)$. Без ограничения общности предположим, что $N=2^M$, M – натуральное число.

Дискретное вейвлет-преобразование последовательности X по отношению к материнскому вейвлету ψ определим [3] как

$$d_{j,\,k}=\sum_{t=0}^{N-1}x_t\psi_{j,\,k}(t). \tag{1}$$

Это преобразование вычисляется для $j=\,=0,\,1,\,\ldots,\,M-1$ и $k=0,\,1,\,2,\,\ldots,\,2^j-1$. Так как имеется N наблюдений, а при вычислении мы получили $N-1$ коэффициент, то недостающий коэффициент обозначим через $d_{-1,\,0}$. Поскольку преобразование (1) является линейным, его можно представить в виде

$$d=\mathrm{W}\cdot X, \tag{2}$$

если предположить, что матрица W является ортогональной, то из (2) следует

$$X=\mathrm{W}'\cdot d, \tag{3}$$

где W' – означает транспонированную матрицу. Соотношение (2) является важным для исследования свойств коэффициентов $\{d_{j,\,k}\}$.

На практике для определения численных значений применяют быстрое вейвлет-преобразование, т. е. пирамидальный алгоритм [3].

Пусть $\psi_{j,\,k}(t)$ – вейвлеты Хаара:

$$\psi^{(Y)}_{j,\,k}(t) = \begin{cases} 2^{-j/2}, & 2^j k \le t < 2^j\left(k+\dfrac{1}{2}\right), \\ -2^{-j/2}, & 2^j\left(k+\dfrac{1}{2}\right) \le t < 2^j(k+1), \\ 0, & \text{иначе.} \end{cases}$$

Если положить $N = 2^3$, то соотношение (3) примет вид

$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \sqrt{2} & 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & \sqrt{2} & 0 & -2 & 0 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & 2 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 2 & -2 & 0 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & 2 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & -2 & 0 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & 2 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} d_{-1,\,0} \\ d_{0,\,0} \\ d_{1,\,0} \\ d_{1,\,0} \\ d_{1,\,1} \\ d_{2,\,0} \\ d_{2,\,1} \\ d_{2,\,3} \end{pmatrix},$$

из которого можно получить $d = \{d_{j,\,k}\}$.

Обозначим через $f(x)$ функцию, ассоциированную с $X = (x_0,\, x_1,\, \ldots,\, x_{N-1})$, т. е. функцию, которая в момент времени t принимает значения $x_t = f(t)$, $t = 0,\, \ldots,\, N-1$. Тогда

$$f(t) = d_{-1,\,0}\varphi(t) + \sum_{j=0}^{2}\sum_{k=0}^{2^j-1} d_{j,\,k}\psi_{j,\,k}(t), \qquad (4)$$

где $\varphi(t) = 1$, $t \in [0,\,1]$ – материнский вейвлет Хаара.

Обычно не рассматривают все уровни разрешения M, а только несколько J, $J < M$, которые соответствуют масштабу 2^{-J}. В этом случае вместо (4) имеет место следующее представление функции $f(t)$:

$$f(t) = \sum_{k=0}^{2^J-1} c_{J,\,k}\varphi_{J,\,k}(t) + \sum_{j=J}^{M-1}\sum_{k=0}^{2^j-1} d_{j,\,k}\psi_{j,\,k}(t). \qquad (5)$$

Из разложения (5) следует, что вейвлет-коэффициентами вейвлет-преобразования последовательности $X = (x_0,\, \ldots,\, X_{N-1})$ является последовательность чисел

$$(c_J,\, d_J,\, d_{J+1},\, \ldots,\, d_{M-1}), \qquad (6)$$

где $c_J = (c_{J,0}, c_{J,1}, ..., c_{J,2^J-1})$, $d_J = (d_{J,0}, d_{J,1}, ..., d_{J,2^J-1})$, $d_{M-1} = \left(d_{M-1,0}, d_{M-1,1}, ..., d_{M-1,2^{M-1}_{-1}}\right)$.

Можно показать, что множеством значений коэффициентов $\{d_{j,\,k}\}$ (1) является

$$\left\{\pm\left(2^{-j/2-1} - i2^{-j/2}\right),\; i = \overline{0,\, 2^{j-1}},\; j = \overline{1,\, M}\right\}.$$

Отсюда следует, что $\mathrm{E}\{d_{j,\,k}\} = 0$, $k = \overline{0,\, 2^{j-1}}$, $j = \overline{1,\, M}$. Ввиду того, что $d_{j,\,k}$ определяется как сумма независимых случайных величин, то в силу центральной предельной теоремы [4]:

$$d_{j,\,k} \xrightarrow{D} N\left(0,\, \Delta_{j,\,k}\right),$$

где $\Delta_{j,\,k} = \mathrm{E}\left\{d^2_{j,\,k}\right\}$.

Из (5) очевидным образом следует, что $d^2_{j,\,k}$ асимптотически эквивалентна случайной величине $\Delta_{j,\,k}\chi^2_1$, где χ^2_1 – случайная величина, имеющая хи-квадрат распределение с одной степенью свободы. Таким образом, для проверки нулевой гипотезы можно применить критерий χ^2, причем нулевая гипотеза применяется, если на всех уровнях разрешения $j = \overline{1,\, M}$ принималась нулевая гипотеза.

Методом статистического моделирования были получены вероятности ошибок первого и второго рода. Для $p \ll 0.5$ вероятность ошибки второго рода была близка к нулю, при $p = 0.5$ вероятность ошибки второго рода приближалась к 0.5. Вероятность ошибки первого рода не превосходила уровня значимости критерия χ^2.

1. Чуи, К. Введение в вейвлеты / К. Чуи // М.: Мир, 2001. – 412 с.
2. Астафьева, Н. М. Вейвлет-анализ: основы теории и примеры применения / Н. М. Астафьева // Успехи физических наук. – 1996. – Т. 11. – С. 11–1170.
3. Chiann, C., Morettin, A. A Wavelet Analysis for time series / C. Chiann, A. Morettin // Journal of Nonparametric Statistics. – 1998. – № 1. – P. 1–46.
4. Ширяев, А. Н. Вероятность / А. Н. Ширяев // М.: Наука, 1980. – 576 с.