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 основе вейвлетов и статистической проверки гипотез рассматривается задача обнаружения моментов измерения свойств дискретных временных рядов.

Введение

Вейвлеты были введены и получили широкое распространение совсем недавно, в конце 1980-х годов как альтернатива классическому взвешенному анализу Фурье. Анализ Фурье основан на разложении сигналов, математической моделью которых являются функции от времени, по базису, состоящему из синусов и косинусов, умноженных на скользящее окно. В вейвлет-анализе окно изначально является осциллирующим и называется материнским вейвлетом. Вместо умножения на синусы и косинусы этот вейвлет произвольно сдвигается и растягивается по временной оси. Таким образом, материнский вейвлет образует другие вейвлеты, которые являются основой вейвлет-анализа.

Анализ Фурье, взвешенный анализ Фурье и вейвлет-анализ основаны на одной и той же концепции. Во всех трех случаях, анализ сигнала заключается в вычислении всех корреляций между заданной функцией и базисными функциями. Синтез производится, как правило, в предположении, что базисные функции образуют ортонормированный базис.

I. Постановка задачи

Рассмотрим проблему обнаружения момента изменения свойств дискретного сигнала как решения следующей задачи. Пусть дана последовательность бит, проиндексированная в каждый момент времени t, т.е.

$$x_t \in \{0,1\},\ t \in N_0.$$

Предположим, что

$$\mathbf{P}\{x_t = 1\} = 1 - \mathbf{P}\{x_t = 0\},\ t < 2^T,$$

$$\mathbf{P}\{x_t = 1\} = 1 - \mathbf{P}\{x_t = 0\},\ t \geq 2^T,$$

Относительно θ и γ не делается никаких предположений, кроме $\theta \neq \gamma$. Требуется оценить параметр T по наблюдениям $\{x_t\}$, $t = 0, 1, 2, \ldots, M$, $M = 2^T$.

II. Дискретное вейвлет-преобразование дискретной последовательности и его асимптотические свойства

Вейвлет-преобразование [1], [2] и статистики, основанные на вейвлет-преобразовании позволяют выявить различные закономерности или же, наоборот, отклонения от заданных закономерностей. Такую возможность предоставляет разделение сигнала на многие уровни с различной степенью детализации.

Пусть $\{x_t\}$, $t = \overline{0, M-1}$, – стационарный временной ряд с нулевым средним, $\gamma(u) = \mathbf{E}\{x_t \cdot x_{t+\tau}\}$, $\tau \in Z$ – автоковариационная функция. Дискретным преобразованием стационарного временного ряда $\{x_t\}$, $t = \overline{0, M-1}$, называется функция переменных $j, k \in Z$, такая, что

$$d_{jk}(x) = \sum_{t=0}^{M-1} x_t \psi_{jk}(t).$$

Очевидно, что

$$\mathbf{E}\{d_{jk}(x)\} = \mathbf{E}\left\{\sum_{t=0}^{M-1} x_t \psi_{jk}(t)\right\} = \\ = \sum_{t=0}^{M-1} \mathbf{E}\{x_t\}\psi_{jk}(t) = 0;$$

$$\mathrm{Var}\{d_{jk}(x)\} = \sum_{t=0}^{M-1}\sum_{t=0}^{M-1} \gamma(t-\tau)\psi_{jk}(t)\psi_{jk}(\tau) = \\ = \sum_{t=0}^{M-1} \gamma(u) \sum_{t=0}^{M-1-|u|} \psi_{jk}(t)\psi_{jk}(t+|u|).$$

В [3] показано, что

$$\mathrm{Var}\{d_{jk}(x)\} \underset{M\to\infty}{\longrightarrow} \eta_{jk}, \eta_{jk},$$

где $\eta_{jk} = \sum_{u=-\infty}^{\infty} \gamma(u)\psi_{jk}(u)$, $j, k \in Z$, –вейвлет-автоковариационная функция.

Используя центральную предельную теорему [4], нетрудно получить, что

$$d_{jk} \xrightarrow{D} N(0, \eta_{jk}).$$

III. Критерий обнаружения момента свойств дискретного временного ряда, основанный на максимуме вейвлет-периодограммы

Для каждого уровня разрешения j периодограмму дискретного временного ряда $\{x_t\}$,

$t = 0, 1, \ldots, M-1$, как квадрат коэффициента вейвлет-преобразования

$$I_{jk}(x) = (d_{jk}(x))^2.$$

Как было показано

$$Z\{d_{jk}\} = N(0, \eta_{jk}),$$

следовательно

$$Z\{I_{jk}(x)\} = \chi_i^2,$$

т.е. распределение вероятностей вейвлет-периодограммы ассимптотически является распределением хи-квадрат с одной степенью свободы.

Математически проблему обнаружения момента изменения свойств дискретного временного ряда можно сформулировать как следующую задачу проверки гипотез. Пусть дана случайная последовательность такая, что первые 2^τ бит распределены по закону Бернулли с параметром θ, а оставшиеся $2^T - 2^\tau$ распределены по закону Бернулли с параметром γ, $\gamma \neq 0$, т.е. пусть дана последовательность $\{\tilde{x}_t\}$, $t = 1, \ldots, 2^T$, такая что

$$Z\{x_t\} = \begin{cases} \mathrm{Bi}(1,0), & t = 1, \ldots, 2^\tau, \quad \tau < T, \\ \mathrm{Bi}(1,\gamma), & t = 2^{\tau+1}, \ldots, 2^T, \end{cases}$$

где $\mathrm{Bi}(\cdot,\cdot)$ – распределение Бернулли. Для простоты преобразуем исходные данные, применяя следующее преобразование $x_t = 2\tilde{x}_t - 1$ к исходной последовательности $\{\tilde{x}_t\}$, $t = 1, \ldots, 2^T$.

В случае использования вейвлета Хаара коэффициенты вейвлет преобразования сигнала x_t сильно упрощаются

$$d_{jk}(x) = \sum_{t=1}^{T} \psi_{jk}(t) x_t =$$
$$= 2^{-\frac{j}{2}} \left(\sum_{t=2^j k}^{2^j(k+\frac{1}{2})-1} x_t - \sum_{t=2^j(k+\frac{1}{2})}^{2^j(k+1)-1} x_t \right),$$

откуда следует независимость вейвлет-коэффициентов $d_{jk}(x)$ на каждом уровне разрешения j, т.к. интервалы, по которым вычисляются вейвлет-коэффициенты не пересекаются.

Основная H_0 и альтернативная H_1 гипотезы формируются так

$$H_0 : Z\{x_t\} = \mathrm{Bi}(1,0), \quad t = 1, \ldots, 2^\tau, \tau < T,$$
$$H_1 : Z\{x_t\} = \mathrm{Bi}(1,\gamma), \quad t = 2^\tau + 1, \ldots, 2^T.$$

Определим максимум вейвлет-периодограммы на каждом уровне разрешения j:

$$I_j = \max_{k=\overline{0,m}} I_{jk}(x), \quad m = 2^{T-(j+1)}.$$

Можно показать, что если $F_{I_j}(x)$ – функция распределения статистики I_j, а $F_{\chi_1^2}(x)$ – функция распределения хи-квадрат с одной степенью свободы, то

$$\left(F_{I_j}(x) - F_{\chi_1^2}(x)\right) \xrightarrow[T\to\infty]{D} 0,$$

тогда P-значение на уровне j для проверки основной гипотезы будет определяться следующим образом [5]

$$P_j = 1 - \left(F_{\chi_1^2}\left(I_j(x)\right)\right)^{T-j}.$$

Решающее правило (критерий) примет вид

$$d = \begin{cases} H_0, & \text{если} \quad P_j > \varepsilon, \, \forall j = \overline{1, T-1}, \\ H_1, & \text{иначе}, \end{cases}$$

здесь ε – уровень значимости критерия.

IV. Результаты компьтерного моделирования

Число моделируемых последовательностей $N = 100$. Для оценки вероятностей ошибок первого рода моделировались бинарные последовательности Бернулли длиной 2^8, 2^{10}, 2^{12} с параметром $\theta = 0.5$. Уровень значимости $\varepsilon = 0.05$. Для оценки вероятностей второго рода моделируемые последовательности состояли из двух однородных фрагментов длиной $2^{T_1} = 2^{T_2} = 2^{T-1}$ с разладкой в момент времени τ, равным 2^7, 2^9, 2^{11}. Первый фрагмент представлял случайную бинарную последовательность с вероятностью появления единицы $\theta = 0.5$, а для второго фрагмента вероятность γ принимала значения 0.05; 0.1; 0.25; 0.30; 0.35; 0.40; 0.45. Для значений $\gamma \ll 0.5$ вероятность ошибки второго рода приближалась к нулю, при $\gamma = 0.5$ вероятность ошибки второго рода приближалась к 0.5. В 96 случаях из 100 момент разладки τ обнаруживался правильно.

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.
5. Ивченко, Г. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев // М.: Высшая школа, 1984.