ТЕСТИРОВАНИЕ СКАЧКОВ ИЗМЕНЕНИЯ ХАРАКТЕРИСТИК СИГНАЛОВ С ПОМОЩЬЮ ВЕЙВЛЕТОВ

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

Применяется вейвлет-анализ для построения тестирующей статистики обнаружения изменений характеристик последовательности независимых случайных величин. Методом статистических испытаний оценивается мощность теста.

Введение

Анализ изменений (в том числе и скачкообразных) характеристик сигнала привлекает внимание многих исследователей. Впервые задача обнаружения "разладки" была сформулирована в 50-е годы XX века в работе [1]. Первоначально эта задача получила применение в промышленном производстве, медицинских исследованиях, геофизике, задачах технической диагностики, обработке сигналов. Таким образом, класс задач обнаружения разладки является очень широким. Эти задачи отличаются одна от другой предположениями о модели наблюдаемого процесса и подходами к ее решению. Существует два основных подхода к решению задачи разладки: методы апостериорного обнаружения и последовательные методы обнаружения. В первом случае предполагается, что в имеющейся последовательности наблюдений в некоторый момент произошло изменение характеристик и на основе полученных наблюдений необходимо оценить момент изменений. При этом свойства полученных оценок изучаются в асимптотической постановке при объеме наблюдений, стремящимся к бесконечности [2]. При последовательном обнаружении на каждом шаге при поступлении нового наблюдения гипотеза о наступлении разладки либо принимается и наблюдения прекращаются, либо отклоняется и наблюдения продолжаются дальше [3].

Одним из основных методов решения задачи обнаружения разладки является метод кумулятивных сумм (CUSUM) [4], причем для изучения CUSUM-статистики используют метод Фурье-анализа [5]. В данной работе применяется вейвлет-анализ для тестирования наличия моментов изменения свойств случайной последовательности случайных величин, которая может служить простейшей математической моделью исследуемого сигнала.

I. Постановка задачи

Пусть наблюдения $x_1, x_2, \ldots, x_n$ являются независимыми случайными величинами, распределенными по нормальному закону распределения вероятностей:

$$x_i \sim N\left(f\left(\frac{i}{n}\right), \sigma^2\right), \qquad (1)$$

где $f(\cdot)$ – известная функция, определенная на отрезке $[0,1]$. В задаче обнаружения момента изменения среднего значения последовательности независимых случайных величин (1) предполагается, что

$$f(t) = \begin{cases} \mu, & 0 \le t < \tau, \\ \mu + \Delta, & \tau < t \le 1, \end{cases} \qquad (2)$$

где τ – неизвестный момент времени. Задача состоит, во-первых, в тестировании факта существования момента изменения среднего значения. Если принимается гипотеза о наличии этого момента, то дальнейшее исследование заключается в оценивании параметров τ, μ, Δ, σ^2. Следует отметить, что ключевой проблемой здесь является оценивание параметра τ. Задача (2) может быть обобщена, если функцию $f(t)$ представить в виде:

$$f(t) = \begin{cases} g(t), & 0 \le t < \tau, \\ h(t), & \tau \le t < 1, \end{cases}$$

где $h(\tau_+) - g(\tau_-)$ – скачок среднего.

II. Вейвлеты и V-тест

Вейвлет-анализ основан на разложении исследуемой функции в ряд по базисным функциям и изучении коэффициентов разложения. Он является удобным аппаратом для исследования разрывов функций, максимумов, минимумов и других особенностей. Удобно использовать объем наблюдений $n = 2^J$, где J – некоторое натуральное число. Вейвлеты представляют собой функции [6]:

$$\left\{\psi_{jk}(t),\, k{=}0,\ldots,2^j-1,\, j{=}0,1,\ldots,J-1\right\}, \qquad (3)$$

где $\psi_{jk}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right)$.

Функция $\psi(t)$ называется материнским вейвлетом. В случае вейвлета Хаара [6]:

$$\psi(t) = \begin{cases} 1, & 0 \le t < 0,5, \\ -1, & 0,5 \le t < 1, \end{cases}$$

Вейвлет-преобразование, примененное к последовательности $\{x_i\}$, $i = 1, 2, \ldots, n$, приводит к рассмотрению вейвлет-коэффициентов

$$d_{jk} = \sum_{i=0}^{n} x_i \psi_{jk}(t_i). \qquad (4)$$

Чтобы изучить вероятность ошибок второго рода используемого статистического критерия, введем последовательность альтернативных гипотез H_{1n} и основную H_0:

H_0: $x_1, \ldots, x_{\tau_n}$ – последовательность независимых случайных величин, распределенных по нормальному закону $N(\mu, 1)$.

H_{1n}: $x_{\tau_n+1}, \ldots, x_n$ – последовательность независимых случайных величин, распределенных по нормальному закону $N(\mu + \Delta_n, 1)$, $\Delta_n = O\left(\sqrt{\log_2 n/n}\right)$.

Поскольку наблюдения x_i независимы и распределены по нормальному закону, то коэффициенты d_{jk}, определенные по формуле (4), также будут независимы и нормально распределенными в силу линейности формулы (4) и ортогональности базиса $\{\psi_{jk}(t)\}$ (3). При условии выполнения нулевой гипотезы H_0 коэффициенты $\{d_{jk}\}$ (4) будут независимы и распределены по закону $N(0, 1)$. При условии выполнения альтернативной гипотезы H_{1n} вейвлет-коэффициенты будут также распределены по нормальному закону, но с ненулевым средним.

Используя эти вероятностные свойства вейвлет-коэффициентов, построим статистику критерия:

$$V = \sum_{j=1}^{J-1} V_j,$$

где $V_j = \sqrt{2^{j-1}} \cdot \sum_{k=0}^{2^j-1} d_{jk}$.

Очевидно, при выполнении гипотезы H_0, статистика критерия $V \sim N(0; \tau \cdot n)$. Таким образом, тест проверки гипотезы H_0 имеет вид:

$$d = \begin{cases} H_0, & \dfrac{|T|}{\sqrt{\tau \cdot n}} < \Delta, \\ H_1, & \dfrac{|T|}{\sqrt{\tau \cdot n}} \ge \Delta, \end{cases}$$

где $\Delta = \Phi^{-1}\left(1 - \dfrac{\varepsilon}{2}\right)$, где ε – уровень значимости V-критерия; $\Phi(t)$ – функция распределения вероятностей стандартного нормального закона. Вероятность ошибки второго рода вычисляется по формуле

$$w_n = \mathrm{P}\left\{ \frac{|T|}{\sqrt{\tau \cdot n}} = \Delta | H_{1n} \right\}.$$

III. Компьютерные эксперименты

Для оценивания мощности теста проводилась серия компьютерных экспериментов. Объемы выборок составляли 16, 32, 64, 128, 256. Предполагается, что из параметров модели $(\tau, \mu, \Delta, \sigma^2)$ неизвестным является только τ. Для каждой выборки размер скачка составлял $\Delta_n = \sqrt{\log_2 n/n}$. Момент наступления скачка определялся по формуле $\tau_n = [n \cdot \lambda]$, где $[a]$ – целая часть числа a. В качестве значений λ брались $0,1$, $0,2$, $0,3$, $0,5$. Результаты моделирования приведены в таблице (уровень значимости $\varepsilon = 0,05$; $\lambda = 0,1$):

n	16	32	64	128	256
w	0,945	0,94	0,9	0,94	0,93

Результаты моделирования показывают эффективность предложенной статистики для обнаружения изменений среднего значения случайной последовательности, поскольку мощность критерия близка к 1.

1. Page, E. S. Continuous inspection schemes / E. S. Page // Biometrika. – 1954. Vol. 41, №. 1. P. 100-115.

2. Page, E. S. Estimating the point of change in a parameter occuring at an unknown point / E. S. Page // Biometrika. – 1955. Vol. 42, №. 4. P. 523-527.

3. Никифоров, И. В. Последовательное обнаружение изменения свойств временных рядов / И. В. Никифоров // М.: Наука. – 1983, 199 с.

4. Gilchrist, C. Note on distribution-free CUSUM technique / C. Mc Gilchrist, K. Woodyer // Technometrics. – 1975. Vol. 17, №. 1. P. 321-325.

5. Lombard, F. Detecting Change Points by Fourier Analysis / F. Lombard // Technometrics. – 1985. Vol. 30, №. 3. P. 305-310.

6. Смоленцев, Н. К. Введение в теорию вейвлетов / Н. К. Смоленцев // М.–Ижевск: НИЦ "Регулярная и хаотическая динамика". – 2010, 292 с.