Известия Гомельского государственного университета
имени Ф. Скорины, №5(50), Ч2, 2008

УДК 517.537.38:519.2

Об одном алгоритме обнаружения особенностей функций с использованием вейвлет–преобразования

М. С. Абрамович, Е. А. Миротин

Введение. Широкое распространение для обнаружения особенностей функций на основе вейвлет-анализа получил алгоритм, изложенный в [2]. Ниже будет указан недостаток этого алгоритма и предложен его вариант с использованием алгоритма, основанного на статистиках от спектральных плотностей. Проведено исследование эффективности модифицированного алгоритма методом статистического моделирования.

Математическая модель. Пусть функция f α-непрерывна по Гёльдеру на [0, 1], за исключением окрестности неизвестной точки $\theta \in (0,1)$. Пусть имеется также выборка наблюдений значений функции f на равномерной сетке при наличии случайного шума:

$$y_i = f\left(i/n\right) + \xi_i, \quad i = 1, \ldots, n, \tag{1}$$

где $\xi_i \sim N(0, \sigma^2)$ – независимые случайные величины, имеющие нормальное распределение с нулевым средним и постоянной дисперсией σ^2. В соответствии с [2] определим дискретное вейвлет-преобразование выборки (1) следующим образом:

$$d_{j,k}^{(\psi)} = \sum_{i=1}^{n} y_i \psi_{j,k}(i), \tag{2}$$

где семейство вейвлетов $\psi_{j,k}$ строится по базисному вейвлету $\psi(i)$ так:

$$\psi_{j,k}(i) = 2^{j/2}\psi(2^j i - k) \ \ (j, k \in Z). \tag{3}$$

Необходимо по выборке (1) оценить момент разладки θ с использованием вейвлет-преобразования (2), (3).

Построение оценки момента разладки. Один из подходов для обнаружения разладок состоит в том, что если рассматриваемая выборка (1) их не содержит, то вейвлет-коэффициенты на каждом уровне разрешения не превосходят заданного порогового значения [5]. Чтобы не сравнивать значения вейвлет-коэффициентов с определенным порогом, будем рассматривать разности соседних коэффициентов на уровнях разрешения и определим величину сдвига k, для которой разность является максимальной. Эта величина сдвига и будет соответствовать моменту разладки. Пусть $N = \left[\alpha\right]$. Поскольку область компактного задания вейвлетов Добеши порядка $D=2N$ – отрезок $[0, D-1]$, рассмотрим сдвиг такого вейвлета, в результате чего получим вейвлет с симметричной областью компактного задания $[-N+1/2, N-1/2]$.

Процедура определения момента разладки состоит из двух этапов: на первом этапе определяется приближенная оценка, и на втором этапе происходит ее уточнение.

Вычисление приближенной оценки момента разладки на **первом этапе** состоит из следующих шагов.

Шаг 1. Вычисляются следующие суммы:

$$\Delta_{j,k}^{(l)} = n^{-1} \sum\nolimits_{l} \psi_{j,k}^{(l)}(i/n) y_i, \tag{4}$$

где $\sum_l$ при $l = 1, 2$ означает суммирование по выборке значений функции (1) с нечетными и четными индексами соответственно (это делается для того, чтобы исключить зависимость статистик (4) между двумя этапами процедуры; параметр m из (6) определяется на втором этапе);

$$\psi^{(1)}_{j,k} = \psi_{j,\tau(k)} - \psi_{j,\tau(k+2)},\ \tau(k) = \begin{cases} 2Nk + N,\ N > 1, \\ k,\ N \le 1, \end{cases} \tag{5}$$

$$\psi^{(2)}_{j,k} = \psi_{j,m-\tau(k)} + \psi_{j,m+\tau(k)}. \tag{6}$$

Шаг 2. Выбирается такой уровень разрешения j_1, чтобы обеспечить достаточную точность оценивания момента разладки и определяется значение величины сдвига k_0:

$$k_0 = \arg\ \max_k \left|\Delta^{(1)}_{j_1,k}\right|,\ k = 0,...,K = (2^{j_1} - 6N)/2N.$$

Шаг 3. Определяется приближенная оценка момента разладки $\theta_0 = \hat{k}_0 / 2^{j_1}$.

На **втором этапе** процедуры полученная оценка момента разладки θ_0 уточняется следующим образом.

Шаг 1. Задается уровень разрешения $j_2 > j_1$. В данном случае k изменяется от *0* до $N_0 = 12N2^{j_2 - j_1}$. Параметр *m* в (6) зависит от оценки момента разладки, найденной на первом этапе, и полагается равным $m = \left[2^{j_2} k_0 - 6N2^{-j_1}\right]$.

Шаг 2. Строятся случайные величины:

$$\eta_k = 1\left\{\left|\Delta^{(2)}_{j_2,k}\right| > Cn^{-1/2}\right\},\ k = 0,...,N_0 - 1, \tag{7}$$

где параметр C рекомендуется выбирать следующим образом [2]:

$$C = \sigma\sqrt{(j_2 - j_1)\ln 2}. \tag{8}$$

Случайные величины $\eta_k,\ k = 0,...,N_0 - 1,$ являются независимыми и имеют распределение Бернулли. Уровень разрешения j_2 с использованием точек $U_0,...,U_k,...,U_{N_0}$ разобьем на отрезки, где отрезок $[U_k, U_{k+1}]$ является областью компактного задания функции $\psi_{j_2,\tau(k)+m}$. Будем полагать, что искомый момент разладки попадает в полуинтервал $\left[U_{k_0}, U_{k_0+1}\right)$. Тогда до момента разладки значение случайной величины η_k будет равно 0, а после – 1.

Шаг 3. Минимизируется следующее выражение:

$$\sum_{k=0}^{\nu}\eta_k + \sum_{k=\nu+1}^{N_0-1}(1-\eta_k). \tag{9}$$

Пусть оно достигает своего минимума при $\nu = k_1$. В качестве уточненной оценки момента разладки принимается $\theta = U_{k_1}$. В случае применения вейвлетов Добеши оценка момента разладки определяется следующим образом:

$$\theta = \frac{-N + 0.5 + \tau(k_1) + m}{2^{j_2}}. \tag{10}$$

Оказывается, оценка момента разладки в алгоритме должна удовлетворять определенному условию, которое не отмечено в [2].

Утверждение. *Для оценки момента разладки (10) при* $n \to \infty$ *имеем* $\theta \le 1/2N$.

Доказательство. 1. Пусть $N = 1$. С учетом (5), (10) последовательно имеем:

$$\theta = (-N + 0.5 + \tau(k_1) + m)/2^{j_2} = (-0.5 + k_1 + m)/2^{j_2} \le (0.5 + k_1 + 2^{j_2}\theta_0 - 62^{-j_1})/2^{j_2} =$$
$$= (0.5 + k_1 + k_0 2^{j_2 - j_1} - 62^{-j_1})/2^{j_2} \le (0.5 + N_0 + K2^{j_2 - j_1} - 62^{-j_1})/2^{j_2} =$$
$$= (0.5 + 12\cdot 2^{j_2 - j_1} + (2^{j_1} - 6)2^{j_2 - j_1}/2 - 62^{-j_1})/2^{j_2} = 0.5/2^{j_2} + 12/2^{j_1} + 0.5 - 3/2^{j_1} -$$
$$- 6/2^{j_2 + j_1} \le 0.5 + 12.5/2^{j_1}.$$

2. Пусть $N > 1$. Тогда аналогично случаю 1) последовательно имеем:

$$\theta=(-N+0.5+\tau(k_1)+m)/2^{j_2}=(-N+0.5+2Nk_1+N+m)/2^{j_2}\le$$
$$\le(1.5+2NN_0+K2^{j_2-j_1}-6N2^{-j_1})/2^{j_2}=(1.5+12N^2 2^{j_2-j_1}+(2^{j_1}-6N)2^{j_2-j_1}/2N-$$
$$-6N2^{-j_1})/2^{j_2}=1.5/2^{j_2}+12N^2/2^{j_1}+1/2N-3/2^{j_1}-6N/2^{j_2+j_1}<1/2N+$$
$$+(12N^2+1.5)/2^{j_1}\le 1/2N+(12N^2+1.5)/2^{j_1}.$$

При $n\to\infty$ имеем $j_1\to\infty$, поэтому правые части выражений, полученные для случаев $N=1$ и $N>1$, не превышают $1/2N$.

Таким образом, из утверждения следует, что значение уточненной оценки момента разладки не превышает $1/2N$. Если предположить, что $\theta<1/2N$, то рассмотренный выше алгоритм позволяет получить оценку момента разладки θ. Нарушение же этого условия не позволит оценить моменты разладок. В связи с этим необходимо провести модификацию алгоритма, заключающуюся в предварительной обработке выборки (1), с целью локализации момента разладки в соответствии с утверждением.

Алгоритм локализации момента разладки. Для локализации момента разладки рассмотрим взвешенную статистику различия непараметрических оценок спектральных плотностей $S_1(\lambda_s)$ и $S_2(\lambda_s)$, вычисленных по первой и второй половинах фрагмента заданной длины M выборки (1) с серединой в точке η:

$$V^2(\eta)=\frac{\sum_{s=1}^{l}\left(S_1(\lambda_s)-S_2(\lambda_s)\right)^2}{\sum_{s=1}^{l}\left(S_1^2(\lambda_s)+S_2^2(\lambda_s)\right)},\qquad(11)$$

где $\{\lambda_1,...\lambda_l\}\in[-\pi,\pi]$ – заданный набор l частот, такой что $\Sigma_{s=1}^{l}(S_1(\lambda_s)-S_2(\lambda_s))^2>0$, $S_1(\lambda_s), S_2(\lambda_s)$ – истинные значения спектральных плотностей.

В [4] найдено распределение статистики (11) и дан такой критерий обнаружения разладки:

$$\text{принимается гипотеза}\quad\begin{cases}H_0,\ \text{åñëè}\ V^2(\eta)<\delta_\varepsilon,\\ H_1,\ \text{åñëè}\ V^2(\eta)\ge\delta_\varepsilon,\end{cases}\qquad(12)$$

где пороговое значение δ_ε для заданного уровня значимости ε приведено в [4]. Кроме того, в [4] доказано, что статистика (11) достигает максимума в момент разладки. На основании критерия (12) и свойства максимума статистики (11) рассмотрим следующий алгоритм обнаружения момента разладки.

Шаг 0. Задается длина скользящего окна *M,* величина сдвига окна d по выборке (1) и число частот l, для которых вычисляются оценки спектральных плотностей.

Шаг 1. Вычисляются оценки спектральных плотностей $S_1(\lambda_s)$, $S_2(\lambda_s)$, $s=1,...,l$ по первой и второй половинам скользящего окна $x_{\eta-u},...,x_\eta,...,x_{\eta+u}$ фиксированного объема $M=2u+1$ (u – параметр) и статистика $V^2(\eta)$ в соответствии с (11).

Шаг 2. Производится смещение скользящего окна по выборке значений последовательности (1) на величину сдвига d и, если скользящее окно не достигло конца выборки, то переход на Шаг 1, иначе на Шаг 3.

Шаг 3. Определяется значение $\eta_{\max}=\arg\max\limits_{\eta=\eta_-,...\eta_+}V^2(\eta)$, где $1<\eta_-<\eta<\eta_+<n$ – некоторые априорно заданные граничные значения.

Если $V^2(\eta_{\max})>\delta_\varepsilon$, то значение $\eta_{\max}$ позволяет локализовать момент разладки θ.

Рассмотрим случай локализации момента разладки для $N = 1$. Разобьем исходную выборку на 4 равные части. С помощью оконного алгоритма локализуем момент разладки θ в окне размера $n/4$. Если момент разладки θ находится в 1-й части, уменьшим выборку в два раза, оставив только 1-ю и 2-ю части. Аналогично поступим, если θ расположен во 2-й или 3-й частях. Если же момент θ расположен в 4-й части, выберем 3-ю и 4-ю части и совершим их инверсию. Очевидно, что такая операция не нарушает предположений относительно исходной выборки. Таким образом, получим новую выборку, в которой момент разладки гарантированно находится в ее первой половине, т.е. выполняется условие $\theta < 1/2N.$ Если же $N > 1$, то можно, взяв разности соответствующего порядка для выборки (1), свести задачу к случаю $N = 1$.

Результаты вычислительных экспериментов. Для исследования эффективности алгоритма из [2] и модифицированного алгоритма генерировались выборки, состоящие из двух фрагментов, описываемых различными полиномами до и после момента разладки θ:

$$y_i = \begin{cases} P_1(i/n) + \xi_i,\ i < \theta, \\ P_2(i/n) + \xi_i,\ i \geq \theta, \end{cases}$$

где $P_1(x) = x^4 + 3x^2 - 2x + 1,\ P_2(x) = x^4 + 3x^2 - 2x - 2,\ \xi_i \sim N(0, \sigma^2)$.

Длина выборки значений функций полагалась равной $n = 1024$. Значение дисперсии $\sigma^2 = 0.2$. Моменты разладок принимали значения как из первой, так и из второй половины выборки. В качестве анализирующего вейвлета использовался вейвлет Добеши порядка $N = 1$. Погрешность вычисления оценки момента разладки определялась по формуле $\delta = (\theta - \theta)/n.$

В следующей таблице приведены результаты тестирования алгоритма из [2] и модифицированного алгоритма. Как видно из таблицы, алгоритм, приведенный в работе [2], практически безошибочно оценивает моменты разладок только тогда, когда они расположены в первой половине выборки. Модифицированный алгоритм одинаково точно определяет моменты разладок вне зависимости от истинного положения момента разладки.

Момент разладки θ	**Алгоритм из [2]**		**Модифицированный алгоритм**	
	Оценка θ	**Погрешность δ**	**Оценка Θ**	**Погрешность δ**
102	99	0.029412	99	0.002930
204	203	0.004902	201	0.002930
307	305	0.006515	307	0.000000
409	409	0.000000	407	0.001953
512	217	0.576172	511	0.000977
614	188	0.416016	615	0.000977
716	252	0.453125	715	0.000977
819	407	0.402344	817	0.001953
921	414	0.495117	921	0.000000

Abstract. The paper considers an algorithm of detecting function discontinuities and their derivatives by means of wavelet transformation and the criterion based on the statistics of values of spectral densities. The results of the numeric experiments are also given.

Литература

1. Дремин, И.М. Вейвлеты и их использование/ И.М.Дремин, О.В.Иванов, В.А.Нечитайло // Успехи физических наук. 2001. Т. 171, № 5. С. 465-501.

2. Raimondo, M. Minimax estimation of sharp change points/ M.Raimondo // The Annals of Statistics. 1998. Vol. 26, № 4. P. 1379-1397.

3.Wang,Y.Jump and sharp cusp detection by wavelets/Y.Wang //Biometrika.1995. №82. P.385-397.

4 Kharin, Yu.S. Detection of spectral change-point in a two-dimensional time series/ Yu.S. Kharin., M.S.Abramovich // Optoelectronics, Instrumentations and Data Processing. 1999. № 2. P. 45-53.

5. Donoho, D. L. Minimax Estimation via Wavelet shrinkage / D. L. Donoho, I. M Johnstone // Annals of Statistics. 1998. Vol. 26, № 3. P. 879-921.

Белорусский государственный университет Поступило 30.04.08