ОБ ОДНОМ АЛГОРИТМЕ ОБНАРУЖЕНИЯ РАЗЛАДОК БИНАРНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТОВ

М.С. Абрамович, М.Н. Мицкевич

Научно-исследовательский институт прикладных проблем математики и информатики
Пр. Независимости, 4, г. Минск, Республика Беларусь
телефон: + (375-17) 2095481; e-mail: abramovichms@bsu.by

Рассмотрен критерий обнаружения разладки бинарных последовательностей, основанный на вейвлет-преобразовании. Предложена процедура построения пороговых значений для тестирования вейвлет-коэффициентов. Проведен анализ эффективности критерия.

Ключевые слова – бинарная последовательность, вейвлет-преобразование, критерий, разладка,

1 ВВЕДЕНИЕ

Для выявления локальных особенностей последовательностей и, в частности, разладок широко применяется вейвлет-анализ [1-3]. Подход, основанный на применении вейвлет-анализа, для обнаружения разладок, заключается в том, что абсолютные значения вейвлет-коэффициентов в силу свойства локальности вейвлет-преобразования имеют максимальные значения в моменты разладок [1, 2].

В работе [3] рассмотрен ряд критериев обнаружения разладок бинарной последовательности, основанных на статистиках от вейвлет-коэффициентов. Со статистической точки зрения проблема обнаружения разладок с использованием вейвлет-коэффициентов может быть также рассмотрена как проблема множественной проверки гипотез относительно вейвлет-коэффициентов.

В настоящей работе предложен критерий обнаружения разладки бинарной последовательности, основанный на проверке гипотез относительно распределения вейвлет-коэффициентов. Методом статистического моделирования получены оценки вероятности ошибок первого и второго рода критерия.

2 МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассмотрим бинарную последовательность независимых случайных величин:

$$X(t)=\{x_t \mid t=0,\dots,T-1\},\ x_t\in\{0,1\}, T=2^M, M\in N. \quad (1)$$

Последовательность (1) имеет разладку в неизвестный момент времени $\tau\in\{\tau_-,\tau_-+1,\dots,\tau_+\}$, где $0<\tau_-<\tau_+<T-1$ - некоторые априорно заданные граничные значения, и состоит из фрагментов бинарных последовательностей:

$$X(t)=\begin{cases}X_1(t), & 0\le t\le\tau-1,\\ X_2(t), & \tau\le t\le T-1,\end{cases}$$

$X_1(t):\ P(x_t=1)=p_1,\ X_2(t): P(x_t=1)=p_2,\ p_1,p_2\in[0,1]$.

Сформулируем нулевую H_0 и альтернативную H_1 гипотезы о распределении вероятностей последовательности (1) следующим образом:

$$H_0: p_1=p_2=0.5,$$
$$H_1: p_1\ne p_2.$$

Определим дискретное вейвлет-преобразование последовательности (1) в соответствии с [2]:

$$d_{j,k}^{(\psi)}=\sum_{t=0}^{T-1}x_t\psi_{j,k}(t),\ j=1,\dots,M\quad k=0,\dots,2^{M-j}-1, \quad (2)$$

где $\psi_{j,k}(t)=2^{-\frac{j}{2}}\psi\left(2^{-j}t-k\right)$, $\psi(t)$ - базисный вейвлет, j - параметр масштаба (уровень разрешения), k - параметр сдвига.

В качестве базисного вейвлета в настоящей работе используется вейвлет Хаара, который в [1] рекомендуется применять для выявления разладок последовательностей:

$$\psi(t)=\begin{cases}1, & 0\le t<0.5,\\ -1, & 0.5\le t<1,\\ 0, & t<0,\ t\ge1.\end{cases}$$

3 КРИТЕРИЙ ОБНАРУЖЕНИЯ РАЗЛАДКИ

Если бинарная последовательность (1) не имеет разладок, то все вейвлет-коэффициенты статистически незначимо отличаются от нуля и по абсолютной величине не превосходят определенного порогового значения. Это пороговое значение будем определять исходя из следующих предположений.

Как показано в [3], если верна гипотеза H_0, то коэффициенты вейвлет-преобразования (2) при $T\to\infty$ имеют асимптотически нормальное стандартное распределение $d_{j,k}^{(\psi)} \xrightarrow{D} N(0,1)$.

Исходные гипотезы H_0 и H_1 о распределении вероятностей последовательности (1) можно заменить набором индивидуальных гипотез H_{0jk} и H_{1jk} о распределении каждого вейвлет-коэффициента:

$$H_{0jk} : d_{j,k}^{(\psi)} \sim N(0,1),$$
$$H_{1jk} : \text{иначе}.$$

Гипотеза H_0 принимается в том случае, если принимаются все гипотезы H_{0jk}, и альтернативная гипотеза H_1 принимается, когда принимается хотя бы одна из гипотез H_{1jk}.

Построим алгоритм тестирования вейвлет-коэффициентов, состоящий из следующих шагов и основанный на подходе из статьи [4].

1. Для каждого вейвлет-коэффициента $d_{j,k}^{(\psi)}$ в предположении, что выполняется нулевая гипотеза H_{0jk}, вычисляется двухстороннее p-значение $p_{jk} = 2(1-\Phi(|d_{j,k}^{(\psi)}|))$, где $\Phi(\cdot)$ – функция стандартного нормального распределения $N(0,1)$.

2. Вычисленные p-значения p_{jk} упорядочиваются по возрастанию $p_{(1)} \le p_{(2)} \ldots \le p_{(m)}$, где каждое значение $p_{(i)}$ соответствует определенному вейвлет-коэффициенту $d_{j,k}^{(\psi)}$, $m = T-1$.

3. В построенном вариационном ряду для всех $i = 1,\ldots,m$ проверяется условие $p_{(i)} \ge (i/m)\alpha$, где α - заданный уровень значимости. Если это условие выполняется, то принимается индивидуальная гипотеза $H_{0(i)}$.

4. Если найдено l, что для всех $i = 1,\ldots,l$ выполняется $p_{(i)} < (i/m)\alpha$, то принимаются индивидуальные гипотезы $H_{1(i)}$ и вычисляется пороговое значение $\lambda_l = \Phi^{-1}(1-p_{(l)}/2)$, с помощью которого можно определить оценку момента разладки.

4 ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ КРИТЕРИЯ

Для оценки эффективности построенного критерия на модельных данных были вычислены оценки вероятности первого и второго рода.

Число моделируемых бинарных последовательностей для оценки вероятности ошибок первого и второго рода предложенного критерия – $K = 1000$.

Для оценки вероятности ошибок первого рода моделировались бинарные последовательности длиной $T = 2^{10}, 2^{11}, 2^{12}, 2^{13}$ с вероятностью $p_1 = 0.5$. Уровень значимости для проверки нулевой гипотезы H_0 полагался равным $\alpha = 0.05$. Полученные оценки вероятности ошибок первого рода для всех анализируемых длин последовательностей не превышали заданный уровень значимости α.

Для оценки вероятности ошибок второго рода моделируемые последовательности состояли из двух однородных фрагментов длиной $T_1 = T_2 = T/2$ с разладкой в моменты $\tau = 2^9, 2^{10}, 2^{11}, 2^{12}$. Первый фрагмент представлял случайную бинарную последовательность с вероятностями $p_1 = p_2 = 0.5$, а для второго фрагмента вероятность p_1 принимала значения $p_1 = \{0.1, 0.2, 0.3, 0.4\}$.

ТАБЛИЦА 1

ОЦЕНКИ ВЕРОЯТНОСТИ ОШИБОК ВТОРОГО РОДА

$T \setminus p_1$	0	0.1	0.2	0.3	0.4
2^{10}	0	0	0	0.45	0.997
2^{11}	0	0	0	0.004	0.97
2^{12}	0	0	0	0	0.517
2^{13}	0	0	0	0	0.002

В таблице 1 приведены оценки вероятности ошибок второго рода предложенного критерия в зависимости от вероятности p_1 и длины последовательности T. Из таблицы 1 видно, что оценки вероятностей ошибок второго рода уменьшаются с ростом длины последовательности T.

ЛИТЕРАТУРА

[1] Wang Y. Jump and sharp cusp detection by wavelets / Y. Wang // Biometrika. – 1995. – Vol. 82. – P. 385-397.

[2] Chiann C. A Wavelet Analysis for Times Series / C. Chiann, P. A. Morettin // Journal of Nonparametric Statistics. – 1998. – № 10. – P. 63–91.

[3] Абрамович М.С. Критерии обнаружения разладок бинарных последовательностей, основанные на вейвлет-преобразовании / М.С. Абрамович, М.Н. Мицкевич // Вестник БГУ. Серия 1: Физ. Мат. Информ. – 2007. – № 2. – С. 78–82.

[4] Simes R.J. An improved Bonferroni procedure for multiple tests of significance / R.J. Simes // Biometrika. – 1986. – Vol. 73. – P. 751-754.