ОБНАРУЖЕНИЕ МОМЕНТОВ ИЗМЕНЕНИЯ ХАРАКТЕРИСТИК СИГНАЛА НА ОСНОВЕ ВЕЙВЛЕТ-АНАЛИЗА

Лобач В. И. Артем А. Ю.
Кафедра математического моделирования и анализа данных, Факультет прикладной математики и информатики, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: lobach@bsu.by, anastasia.artyomyandex.ru

Рассматривается задача обнаружения момента скачкообразного изменения математического ожидания случайной последовательности на основе критериальной статистики, построенной по коэффициентам вейвлет-преобразования исходного сигнала.

Теория вейвлетов дает удобный и эффективный инструмент для решения многих практических задач. Основная область применения вейвлет-преобразований – анализ и обработка сигналов, изображений, функций, нестационарных во времени или неоднородных в пространстве, когда результаты анализа должны содержать не только частотную характеристику сигнала (распределение энергии сигнала по частотным составляющим), но и сведения о локальных координатах, на которых проявляют себя те или иные группы частотных составляющих или на которых происходят быстрые изменения частотных составляющих сигнала. По сравнениям с расположением сигналов на ряды Фурье вевлеты способны с гораздо более высокой точностью представлять локальные особенности сигналов, вплоть до разрывов 1-го рода (скачков). В отличие от преобразования Фурье, вевлет-преобразование одномерных сигналов обеспечивает двумерную развертку, при этом частота и координаты рассматриваются как независимые переменные, что дает возможность анализа сигналов сразу в двух пространствах. Одна из главных и особенно плодотворных идей вейвлетного представления сигналов на различных уровнях декомпозиции заключается в разделении функций приближения к сигналу на две группы: аппроксимирующую – грубую, с достаточно медленной временной динамикой изменений, и детализирующую – с локальной и быстрой динамикой изменений на фоне плавной динамики, с последующим их дроблением и и детализацией на других уровнях декомпозиции сигналов. Скачкообразное изменение математического ожидания описывается разрывом 1-го рода непрерывной функции, поэтому вейвлет-коэффициенты разложения сигнала по вейвлет-базису в момент разладки имеют максимальные абсолютные значения. На этом свойстве вейвлет-коэффициентов строятся статистики критериев обнаружения моментов изменения свойств исследуемого сигнала.

I. Вейвлет-анализ временных рядов

Рассмотрим цифровой сигнал длительности $T = 2^M$.

$$x = (x_0, x_1, \dots, x_{T-1}),\ x_t \in R. \tag{1}$$

Будем рассматривать дискретное вейвлет-преобразование этого временного ряда (1). Оно заключается в вычислении коэффициентов

$$g_k = \sum_{t=0}^{T-1} x_t \psi_{jk}(t),\ j = \overline{1, M},\ k = \overline{0, 2^{M-j}-1},$$

где $\{\psi_{jk}\}$ – семейство вейвлетов, полученное из материнского вейвлета по формуле

$$\psi_{jk}(t) = \sqrt{2^j}\psi(2^j t - k),\ j = \overline{1, M},\ k = \overline{0, 2^{M-j}-1},$$

где j – параметр масштаба (так называемый уровень разрешения), k – параметр сдвига. Часто в качестве материнского вейвлета берут вейвлет Хаара:

$$\psi(t) = \begin{cases} 1, & 0 \le t < \frac{1}{2}, \\ -1, & \frac{1}{2} \le t < 1, \\ 0, & t < 0 <, t \ge 1. \end{cases}$$

Для анализа сигналов произвольной длительности T (не обязательно степенью двойки $T = 2^M$) можно поступать следующим образом.

1. Использовать непрерывное вейвлет-преобразование, но тогда значительно возрастает вычислительная сложность алгоритма.
2. Использовать диадное вейвлет-преобразование в “в скользящем окне” длины степени двойки.
3. Дополнить временной ряд до длины, равной степени двойки, добавляя в конец ряда значения, полученные экстраполяцией исходного ряда.

Важным приложением вейвлет-анализа временных рядов является обнаружение моментов “разладки” сигналов.

II. Обнаружение “скачкообразных” изменений среднего значения сигнала

Обнаружение “скачкообразных“ изменений среднего значения сигнала, которые описывают процессы в экономике, технике и других приложениях, является актуальной прикладной задачей. Предположим, что сигнал описывается следующей моделью:

$$x(t) = f(t)+\varepsilon(t),\ f(t) = \begin{cases} \mu, & 0 \le t \le t_0 - 1, \\ \mu + \tau, & t \ge t_0, \end{cases}$$

где $\varepsilon(t)$, $t = 0, \dots, T-1$ – независимые одинаково распределенные случайные величины с нулевыми математическими ожиданиями и конечной дисперсией σ^2; t_0 – момент времени, в который происходит скачкообразное изменение математического ожидания (среднего) временного ряда, τ – величина скачка.

Определим нулевую гипотезу H_0, состоящую в том, что $\tau = 0$. т.е. анализируемый сигнал не имеет “скачкообразных” изменений среднего, и альтернативную гипотезу $H_1 = \bar{H}_0$, состоящую в том, что $\tau \neq 0$, т.е. в момент времени t_0 временной ряд имеет “скачкообразное” изменение.

III. Критерий, основанный на максимуме сумм вейвлет-коэффициентов

Определим следующую статистику:

$$V_j = \sqrt{2^M} \sum_{k=0}^{2^{M-j}} c_{jk}.$$

Так как вейвлет-коэффициенты c_{jk} на каждом уровне разрешения являются независимыми одинаково распределенными случайными величинами. т.е. удовлетворяют условиям центральной предельной теоремы, то можно показать, что при выполнении нулевой гипотезы H_0 статистика $\dfrac{V_j}{\hat{\sigma}\sqrt{T}}$ имеет асимптотически стандартное нормальное распределение $N(0,1)$. Пусть

$$V = \max\{|V_0|, |V_1|, \dots, |V_M|_1\}.$$

Квантиль уровня n функции распределения случайной величины V вычисляется по формуле

$$Q(n) = -\sqrt{n}\Phi^{-1}\left(\left(1 - n^{\frac{1}{M}}\right)/2\right),$$

поэтому критерий проверки нулевой гипотезы об отсутствии скачкообразных изменений среднего значения сигнала может быть записан в виде

$$d = \begin{cases} H_0, & V > -\sqrt{n}\Phi^{-1}\left(\dfrac{1-(1-\alpha)^{\frac{1}{M}}}{2}\right), \\ H_1, & \text{иначе}. \end{cases}$$

IV. Обнаружение разладок бинарных последовательностей

Рассмотрим частный случай задачи обнаружения момента изменения свойств цифрового сигнала, имеющий важное значение в криптографии, а именно задачу обнаружения момента разладки бинарной последовательности. Будем рассматривать сигнал x_t, $t = 0, 1, \dots, T-1$, значения которого либо $+1$, либо -1, т.е. $x_t \in \{-1, 1\}$. Предположим, что этот сигнал имеет разладку в некоторый неизвестный момент времени τ, $\tau \in \{\tau_-, \tau_- + 1, \dots, \tau_+\}$, где $0 < \tau_- < \tau_+ < T-1$, – некоторые априорно заданные граничные значения. Таким образом, последовательность $\{x_t\}$, $t = 0, 1, \dots, T-1$, состоит из двух фрагментов бинарных последовательностей:

$$x_t = \begin{cases} x_{1t}, & 0 \le t \le \tau - 1, \\ x_{2t}, & \tau \le t \le T-1, \end{cases}$$

причем $\mathrm{P}\{x_{1t} = 1\} = p_1$, $\mathrm{P}\{x_{2t} = 1\} = p_2$, $p, p_1 \in (0,1)$.

Нулевая гипотеза состоит в том, что последовательность $\{x_t\}$, $t = 0, 1, \dots, T-1$, не имеет разладки, альтернатива в том, что разладка произошла в некоторый неизвестный момент времени τ:

$$H_0: \quad p_1 = P_2,$$
$$H_1, \quad p_1 \neq P_2.$$

Критерий обнаружения разладки, основанный на использовании порогового значения, состоит в следующем: последовательность $\{x_t\}$, $t = 0, 1, \dots, T-1$ имеет разладку, если хотя бы на одном уровне размещения j выполняется условие

$$|c_{jk}| > \lambda,\ j = \overline{1, M},\ k = \overline{0, 2^{M-j} - 1}.$$

В качестве порогового значения будем использовать универсальное пороговое значение для всех уровней разрешения, предложенное в работе []:

$$\lambda = \sigma\sqrt{2\log_2 T},$$

где σ – среднеквадратическое отклонение элементов анализируемой последовательности.

Если факт наличия разладки установлен, то оценка момента разладки вычисляется по формуле

$$\hat{\tau} = 2^{j*}(k* + 0.5),$$

где $(j*, k*) = \mathrm{argmax}|c_{jk}|$.

V. Компьютерные эксперименты

Для анализа эффективности описанного выше алгоритма обнаружения момента разладки разработана программа на языке $C\sharp$ в среде Microsoft Visual Studio 2010. Проводилась оценка вероятностей ошибок первого и второго рода. Число моделируемых последовательностей 100. Для оценки вероятностей ошибок первого рода моделировались бинарные последовательности длиной $T = 2^8, 2^{10}, 2^{12}$ с вероятностью $p_1 = 0.5$. Для оценки вероятностей ошибок второго рода моделировались последовательности длины $T = 2^{14}$ с разладками в моменты $\tau = 2^7, 2^9, 2^{11}$, $p_1 = 0.5$; $p_2 \in \{0.1, 0.2, 0.3, 0.4\}$. Результаты проведенных экспериментов показали работоспособность и эффективность предложенного критерия обнаружения разладки бинарных последовательностей.