ПРИМЕНЕНИЕ ВЕЙВЛЕТОВ В СТАТИСТИЧЕСКОМ АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ

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

В работе представлен алгоритм прогнозирования временных рядов, основанный на разложении фильтрационной и прогнозной плотностей по вейвлет-базису Хаара.

Введение

Термин «вейвлет» появился в середине 80-х годов прошлого века в связи с анализом свойств сейсмических и акустических сигналов. В настоящее время вейвлеты широко применяются в задачах распознавания образов, при обработке и синтезе различных сигналов, при анализе изображений и в других областях исследований.

Вейвлет-преобразование одномерного сигнала состоит в его разложении по базису, полученному с помощью сдвигов и растяжений (сжатий) одной функции, называемой материнским вевлетом. Каждая их этих функций этого базиса характеризует как определенную частоту, так и ее локализацию во времени. Таким образом, в отличие от традиционно применяемого для анализа сигналов преобразования Фурье вейвлет-преобразование обеспечивает двумерную развертку исследуемого одномерного сигнала, при этом частота и координаты рассматриваются как независимые переменные.

Вейвлеты можно применять для обработки изображений и статистического анализа данных. При анализе временных рядов одной из проблем является его разложение на компоненты: тренд, сезонные компоненты и неизвестный период.

Другой важной проблемой анализа временных рядов является прогнозирование будущих значений временного ряда. Одним из новых методов прогнозирования является вейвлет-анализ.

Пусть имеется фрагмент $\{x_1, x_2, \ldots, x_t\}$ временного ряда с нулевым средним. Из теоремы Парсеваля [1] для дискретного Фурье-преобразования следует, что выборочная дисперсия временного ряда $\{x_t\}$ равна:

$$S^2 = \frac{1}{T}\sum_{t=1}^{T} x_t^2 = \sum_{j=1}^{T} |d(\omega_j)|^2, \qquad (1)$$

где $d(\omega) = \sum\limits_{t=1}^{T} x_t^2 \exp\{-i\omega t\}$ – дискретное преобразование Фурье ряда $\{x_1, \ldots, x_T\}$, $\omega_j = \frac{2\pi j}{T}$, $j = 0, 1, \ldots, [\frac{n}{2}]$, – частота Фурье.

Равенство (1) свидетельствует о том, что выборочная дисперсия временного ряда может быть сжата в сумму, полученную с помощью множества частот. Выражение $|d(\omega_j)|^2$ как функция от ω_j является графиком частотной функции ряда. Эта функция является оценкой реального спектра $f(\omega)$ процесса. Этот факт дает возможность взглянуть на ряд в частной области определения вместо временной области определения. Если спектр временного ряда достигает максимума на частоте ω_0, тогда можно сделать вывод, что в разложении Фурье временного ряда компонента с частотой ω_0 оценивает большую часть выборочной дисперсии ряда. Еще большие возможности для статистического анализа временных рядов предоставляет вейвлет-анализ.

I. Постановка задачи

В наиболее общей постановке нелинейная негауссовская модель в пространстве состояний временного ряда может быть представлена следующим образом [2]:

$$y_t \sim p(y_t|x_t),\; x_{t+1} \sim p(x_{t+1}|x_t),\; x_1 \sim p(x_1) \qquad (2)$$

для $t = 1, \ldots, T$. Предполагается, что

$$p(Y_T|\alpha) = \prod_{t=1}^{T} p(y_t|\alpha_t),\; p(x) =$$

$$= p(x_1) \prod_{t=1}^{T-1} p(x_{t+1}|x_t), \qquad (3)$$

где $Y_T = (y_1', y_2', \ldots, y_T')'$, $x = (x_1', x_2', \ldots, x_T')'$. Плотность распределения вероятностей $p(y_t|x_t)$ определяет статистическую зависимость между вектором наблюдений y_t и вектором состояний x_t. Плотность распределения вероятностей $p(x_{t+1}|x_t)$ определяет зависимость между вектором состояний в будущий момент времени $t+1$ и вектором состояний в текущий момент времени t. Если зависимости между y_t и x_t, а также между x_{t+1} и x_t являются линейными, то модель (2), (3) называется линейной негауссовской моделью в пространстве состояний. Если все плотности распределения вероятностей $p(y_t|x_t)$, $p(x_{t+1}|x_t)$ являются гауссовскими, но, по крайней мере, одна из зависимостей между y_t и x_t, x_{t+1} и x_t является нелинейной, то модель (2), (3) называется нелинейной гауссовской моделью в пространстве состояний. В случае линейной гауссовской модели в пространстве состояний основным инструментом исследования временных рядов является фильтр Калмана [3]. Требуется по наблюдениям $Y_T = (y_1', \ldots, y_T')'$ оценить будущие значения x_{T+1} y_{T+1} временного ряда (x_t, y_t) при условии, что плотности распределения вероятностей $p(y_t|x_t)$, $p(x_{t_1}|x_t)$ известны.

Хорошо известно [4], что оптимальными в среднеквадратическом смысле оценками x_{t+1} и y_{t+1} по наблюдениям Y_t являются условные математические ожидания

$$\hat{x}_{t+1} = E\{x_{t+1}|Y_t\},\ \hat{y}_{t+1} = E\{y_{t+1}|Y_t\} \qquad (4)$$

II. Фильтрационная и прогнозная плотности распределения вероятностей

Для вычисления оценок по формулам (4) необходимо знать фильтрационную плотность распределения вероятностей $p(x_t|Y_t)$ и прогнозную плотность распределения вероятностей $p(x_{t+1}|Y_t)$. Получим рекуррентные формулы для вычисления этих условных плотностей распределения вероятностей. Согласно формуле Байеса [3, 4] имеем соотношение для фильтрационной плотности

$$p(x_{t+1}|Y_{t+1}) = \frac{p(x_{t+1},\, y_{t+1}|Y_t)}{p(y_{t+1}|Y_t)}. \qquad (5)$$

Далее

$$p(x_t,\, x_{t+1},\, y_{t+1}|Y_t) =$$
$$= p(x_t|Y_t) \cdot p(x_{t=1}|x_t) \cdot p(y_{t+1}|x_{t+1}). \qquad (6)$$

Интегрируя (6) по x_t, получим

$$p(x_{t+1},\, y_{t+1}|Y_t) =$$
$$= \int p(x_t|Y_t) \cdot p(x_{t+1}|x_t) \cdot p(y_{t+1}|x_{t+1}) dx_t. \qquad (7)$$

Формула (7) дает выражение для прогнозной плотности. Из (5) и (7) получим искомую рекуррентную формулу для фильтрационной плотности

$$p(x_{t+1}|Y_{t+1}) =$$
$$= \frac{\int p(x_t|Y_t) \cdot p(x_{t+1}|x_t) \cdot p(y_{t+1}|x_{t+1}) dx_t}{\int\int p(x_t|Y_t) \cdot p(x_{t+1}|x_t) \cdot p(y_{t+1}|y_t) dx_t dx_{t+1}}, \qquad (8)$$

$t = 1, 2, \ldots, T$.

III. Алгоритм прогнозирования

Предлагается алгоритм прогнозирования, основанный на вычислении оценок (4) с помощью фильтрационной и прогнозной плотностей (7), (8), которые представлены в виде разложения по веввлет-базису Хаара. Приведем процедуру приближения функции $f(x)$, которая применяется для приближения фильтрационной и прогнозной плотности:

1. Задается массив значений функции $f(x)$ в точках $x_k = k \cdot \Delta$, $k = \overline{0,\, N-1}$, $N = 2^M$.

$$f = \{f_0,\, f_1,\, \ldots,\, f_{N-1}\}.$$

2. Строится более грубое приближение f^1 последовательности f:

$$f^1 = \{f_k^1\}_{k=0}^{\frac{N}{2}-1} =$$
$$= \left\{ \frac{f_0 + f_1}{2},\, \frac{f_2 + f_3}{2},\, \ldots,\, \frac{f_{N-2} + f_{N-1}}{2} \right\}.$$

Длина последовательности f^1 будет составлять половину длины исходной последовательности.

3. Детализация функции $f(x)$:

$$d^1(x) = f(x) - f^1(x).$$

С помощью вейвлета Хаара получим следующее выражение для функции $d^1(x)$:

$$d^1(x) = \sum_{k=0}^{\frac{N}{2}-1} d_k^1 \psi^H(\frac{x}{2} - k),$$

где $\{d_k^1\}_{k=0}^{\frac{N}{2}-1} = \left\{ \frac{f_0-f_1}{2},\, \frac{f_2-f_3}{2},\, \ldots,\, \frac{f_{N-2}-f_{N-1}}{2} \right\}$ – детализирующие коэффициенты.

4. Функция $f(x)$ восстанавливается следующим образом:

$$f(x) = f^1(x) + d^1(x).$$

Отметим, что данную процедуру можно повторить с использованием функции $f^1(x)$ вместо $f(x)$ и т.д. В результате имеем разложение функции $f(x)$ по вейвлет-базису Хаара:

$$f(x) = f^M(x) + \sum_{j=1}^{M} d^j(x).$$

Проведенные компьютерные эксперименты показывают эффективность данного подхода к прогнозированию будущих значений временного ряда $(x_t,\, y_t)$.

1. Чуп, К. Введение в вейвлеты / К. Чуп // М.: Мир, 2012. – 412 с.
2. Durbin, J. Time Series Analysis by State Space Methods / J. Durbin, S. J. Koopman // Oxford University Press, 2012. – 346 p.
3. Липцер, Р. / Статистика случайных процессов / Р. Липцер, А. Ширяев // М.: Наука, 1974. – 696 с.
4. Ивченко, Г. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев // М.: Высшая школа, 1984. – 286 с.