ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ МАРКОВСКОГО ТИПА НА ОСНОВЕ ВЕЙВЛЕТ-АНАЛИЗА

С.В. Лобач, Е.Е. Жук
Кафедра математического моделирования и анализа данных, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: sergey.vi.@gmail.com, zhukee@mail.ru

Для решения задачи прогнозирования временных рядов, являющихся марковскими процессами, предлагается подход, основанный на формуле Байеса и применении вейвлетов Хаара для вычисления условных математических ожиданий, которые представляют собой оптимальные прогнозирующие статистики. Проведенные модельные расчеты демонстрируют эффективность данного подхода.

Значительный круг задач статистики случайных процессов можно сформулировать следующим образом [1]. На некотором вероятностном пространстве $(\Omega,\,\mathcal{F},\,\mathbf{P})$ задан частично наблюдаемый случайный процесс $(x_t,\,y_t)$, $t\in T$, где T может быть конечным интервалом $T=[0,\,L]$, либо бесконечным интервалом $T=[0,\,+\infty)$, либо дискретным множеством $T=\{0,\,1,\,2,\,\ldots\}$. В последнем случае случайный процесс $(x_t,\,y_t)$ называют временным рядом. Предположим, что у временного ряда наблюдается лишь вторая компонента (y_t), $t\geq 0$. В каждый момент времени t требуется, основываясь на наблюдениях $\bar{y}_t = \{y_s,\,0\leq s\leq t\}$, оценить ненаблюдаемое значение y_τ. Если $\tau=t$, то такая задача оценивания называется задачей фильтрации, если $\tau>t$, то задачей прогнозирования (экстрополяции), если $\tau<t$, то задачей сглаживания (интерполяции).

Первым рассмотрел задачу прогнозирования в такой постановке Винер. Для стационарного двумерного процесса с бесконечным временем была построена оптимальная в среднеквадратическом смысле оценка $\hat{x}_\tau$ в виде

$$\hat{x}_\tau = \int\limits_0^t \mathrm{W}(t,\,s)y_s ds,$$

где весовая функция $\mathrm{W}(t,\,s)$ является решением интегрального уравнения Винера – Хопфа [2]. Колмогоров [3] рассмотрел частный случай подхода Винера, когда $y_t=x_t$, т. е. имеется одномерный стационарный процесс (x_t), $t\geq 0$, с дискретным временем. Прогнозирующая статистика строилась в виде

$$\hat{x}_{t+m} = a_0x_t + a_1x_{t-1} + \ldots + a_nx_{t-n},$$

где коэффициенты $\{a_0, a_1, \ldots, a_n\}$ находились из условия минимизации специально построенной квадратичной формы, m – горизонт прогнозирования.

После этих результатов Винера и Колмогорова появилось большое количество работ по этой тематике, рассматривающих задачу прогнозирования временных рядов при различных предположениях относительно свойств временного ряда. Например, проблема робастного прогнозирования рассматривалась в работе [4].

В данной работе для решения задачи прогнозирования предлагается использовать подход, основанный на формуле Байеса и применении вейвлет-анализа к возникающим численным расчетам.

Хорошо известно [5], что если $\mathrm{E}\left\{x_t^2\right\}<\infty$, то оптимальной в среднеквадратическом смысле оценкой x_τ по $\bar{y}_t = \{y_0,\,y_1,\,\ldots,\,y_t\}$ является условное (апостериорное) среднее $\hat{x}_\tau = \mathrm{E}\left\{x_\tau|\bar{y}_t\right\}$. Таким образом, решение задачи прогнозирования сводится к нахождению условных математических ожиданий $\mathrm{E}\left\{x_\tau|\bar{y}_t\right\}$. Условные математические ожидания $\mathrm{E}\left\{x_\tau|\bar{y}_t\right\}$ можно вычислить по формуле Байеса. Выражения, полученные с помощью формулы Байеса являются очень громоздкими, что затрудняет их практические использование. Однако свойство марковости временного ряда и применение вейвлет-анализа позволяют упростить вычисления.

Рассмотрим частично наблюдаемый двумерный временной ряд $(x_k,\,y_k)$, $k=0,\,1,\,2,\,\ldots$, $x_k\in R$, $y_k\in R$, определяемый соотношениями

$$x_{k+1} = f\left(x_k,\,k\right) + \xi_k, \tag{1}$$

$$y_k = h\left(x_k\right) + \eta_k, \tag{2}$$

где $f(\cdot)$ и $h(\cdot)$ – известные функции $(\xi_k,\,\eta_k)$, $k = 0,\,1,\,2,\,\ldots$ – последовательность взаимно независимых случайных величин. Известны плотности распределения вероятностей случайных величин $\omega_k\sim p_1(x)$, $v_k\sim p_2(x)$. Плотность распределения вероятностей случайной величины $x_0\sim p_0(x)$. При этих условиях временной ряд $(x_t,\,y_t)$, $t=0,\,1,\,2,\,\ldots$, определяемый (1), (2), является марковским процессом.

Требуется по наблюдениям $\{y_k,\,k=0,\,1,\,\ldots,\,T\} = \bar{y}_T$ построить оценки значений x_{T+m}, y_{T+m}, т. е. найти условные математические ожидания

$$\hat{x}_{T+m} = \mathrm{E}\left\{x_{T+m}|\bar{y}_t\right\},\ \hat{y}_{T+m} = \mathrm{E}\left\{y_{T+m}|\bar{y}_t\right\}. \tag{3}$$

Как известно, конечномерные плотности распределения вероятностей марковского процесса $p\left(\bar{x}_k,\, \bar{y}_k\right)$ полностью определяются начальной плотностью $p\left(x_0,\, y_0\right)$ и условными плотностями $p\left(x_k,\, y_k | x_{k-1},\, y_{k-1}\right)$ – плотностью распределения вероятностей перехода за один шаг из состояния $\left(x_{k-1},\, y_{k-1}\right)$ в состояние $\left(x_k,\, y_k\right)$ марковского случайного процесса $\left(x_t,\, y_t\right)$, $t \geq 0$:

$$p\left(\bar{x}_k,\, \bar{y}_k\right) = \prod_{i=1}^{k} p\left(x_i,\, y_i | x_{i-1},\, y_{i-1}\right) \cdot p\left(x_0,\, y_0\right).$$

Введем следующие обозначения:

$$\omega\left(\bar{x}_m | \bar{y}_m\right) = \frac{\omega\left(\bar{x}_m,\, \bar{y}_m\right)}{\omega\left(\bar{y}_m\right)} - \tag{4}$$

условная плотность распределения вероятностей значений $\bar{x}_m$ при условии, что наблюдается реализация $\bar{y}_m$.

$$\omega_{ps}\left(x_m;\, m\right) = \omega\left(x_m | \bar{y}_m\right) = \frac{\omega\left(x_m,\, \bar{y}_m\right)}{\omega\left(\bar{y}_m\right)} - \tag{5}$$

условная плотность распределения вероятностей значения x_m при условии, что наблюдается реализация $\bar{y}_m$.

Имеет место формула Байеса для условной плотности распределения вероятностей $\omega_{ps}\left(x_{m+1};\, m+1\right)$:

$$\omega_{ps}\left(x_{m+1};\, m+1\right) = \tag{6}$$

$$= \frac{\sum\limits_{x_m} p\left(x_{m+1},\, y_{m+1} | x_m,\, y_m\right) \cdot \omega_{ps}\left(x_m;\, m\right)}{\sum\limits_{x_m} \sum\limits_{x_{m+1}} p\left(x_{m+1},\, y_{m+1} | x_m,\, y_m\right) \cdot \omega_{ps}\left(x_m;\, m\right)}.$$

Используя формулы (4)-(6), получим последовательность плотностей распределения вероятностей $p_0(x)$, $\omega_{ps}\left(x_1;\, 0\right)$, $\omega_{ps}(x_1;\, 1)$, $\ldots$, $\omega_{p_s}\left(x_{m+2};\, m\right)$, $\ldots$, $\omega_{ps}\left(x_{m+T};\, m\right)$.

Оптимальные в среднеквадратическом смысле прогнозирующие статистики (3) имеют вид

$$\hat{x}_{m+T} = \int\limits_{-\infty}^{\infty} x_{m+T}\, \omega_{ps}\left(x_{m+T};\, m\right) dx_{m+T},$$
$$\hat{y}_{m+T} = h\left(\hat{x}_{m+T}\right).$$

Для вычисления $\omega_{ps}\left(x_m;\, m\right)$, $\omega_{ps}\left(x_{m+1};\, m\right)$, $\ldots$, $\omega_{ps}\left(x_{m+T};\, m\right)$ используется система вейвлетов Хаара

$$\psi_{j,\,k}^{(H)}(t) = \begin{cases} 2^{-j/2}; & 2^j k \leq t < 2^j \left(k + \dfrac{1}{2}\right); \\ -2^{-j/2}; & 2^j \left(k + \dfrac{1}{2}\right) \leq t < 2^j (k+1). \end{cases}$$

Проведено численное моделирование для случая, когда функция $f(x)$ – линейная, $h(x)$ – квадратичная. Результаты моделирования продемонстрировали эффективность предложенного метода прогнозирования временных рядов.

1. Липцер, Р. Ш., Ширяев, А. Н. Статистика случайных процессов / Р. Ш. Липцер, А. Н. Ширяев // М.: Наука, 1974. – 696 с.
2. Winer, N. Extrapolation, interpolation and smoothing of stationary time series / N. Winer // N.Y.: J. Wiley&Sons, 1949. – 236 p.
3. Колмогоров, А. Н. Интерполирование и экстраполирование стационарных случайных последовательностей / А. Н. Колмогоров // Известия АН СССР, сер. матем. – 1941. – № 5. – С. 5–41.
4. Харин, Ю. С. Оптимальность и робастность в статистическом прогнозировании / Ю. С. Харин – Минск: БГУ, 2008. – 263 с.
5. Ивченко, Г. И., Медведев Ю. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев // М.: Наука, 1984. – 248 с.