О ПРОГНОЗИРОВАНИИ ВРЕМЕННЫХ РЯДОВ В ФОРМЕ МОДЕЛЕЙ В ПРОСТРАНСТВЕ СОСТОЯНИЙ

Е. Е. Жук, С. В. Лобач
Кафедра математического моделирования и анализа данных,
Факультет прикладной математики и информатики,
Белорусский государственный университет
Минск, Республика Беларусь
E-mail: iserega88@mail.ru, zhukee@mail.ru

Обсуждается один из подходов к прогнозированию временных рядов, основанный на рассмотрении будущих значений временного ряда как пропущенных наблюдений. Приводятся рекуррентные формулы для расчета прогнозных значений временного ряда. Результаты компьютерного моделирования приведены для AR (2) модели временного ряда.

Введение

Модели в пространстве состояний были предложены в середине прошлого века для математического описания систем в теории оптимального управления и автоматического регулирования [1]. Математическая теория оптимального управления позволяет свести оптимального управления и процесс построения оптимального управления к решению краевой задачи для дифференциальных уравнений (обыкновенных либо в частных производных). Результаты, полученные А. М. Летовым [2] и Р. Калманом [1] явились основой направлений синтеза систем оптимальной стабилизации: аналитического конструирования регуляторов при измеряемом векторе состояния системы и оптимального управления при неполной информации [5]. Настоящее время модели в пространстве состояний применяются во многих сферах научных исследований. В частности, все известные параметрические модели временных рядов , такие как AR(n), ARMA(n, m), ARIMA(n, d, m), ARCH(p) и т.д., могут быть представлены в виде векторных линейных моделей в пространстве состояний.

I. Линейные гауссовские модели в пространстве состояний и фильтр Калмана

Рассмотрим линейную гауссовскую модель в пространстве состояний

$$x_{t+1} = F_t x_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, Q_t), \qquad (1)$$

$$y_t = H_t x_t + \eta_t, \quad \eta_t \sim N(0, R_t), \qquad (2)$$

$$x_0 \sim N(a_0, P_0), \quad t = 1, 2, \ldots, T.$$

Матрицы F_t, H_t, R_t, Q_t размеров $(n{\times}n)$, $(m{\times}n)$, $(n{\times}n)$, $(m{\times}n)$ предполагаются известными. Изначально начальные значения параметров a_0, P_0 считаются известными, хотя нетрудно рассмотреть случай, когда они неизвестны. m-вектор y_t называется наблюдением, ненаблюдаемый n-вектор x_t называется состоянием. Случайные векторы (ε_t, η_t), $t{=}1, ..., T$, являются последовательностью независимых гауссовских случайных векторов с нулевыми математическими ожиданиями и с известными ковариационными матрицами Q_t и R_t.

Уравнение (2) является стандартной моделью многомерной множественной регрессии, вектор коэффициентов которой x_t изменяется во времени; изменение во времени x_t определяется векторной авторегрессионной моделью первого порядка VAR(1), описывемой уравнением (1). Векторный процесс (x_t, y_t), $t{=}1, \ldots, T$, является марковским, что позволяет привлекать аппарат марковских процессов для исследования его свойств.

Введем обозначения $Y_t{=}\{y_1, ..., y_t\}$, $t{=}1, ..., T$, и сформулируем основные проблемы, связанные с моделью (1)–(2).

1. Проблема фильтрации. Вычисление рекуррентных оценок $\mu_{t|t} = \mathbf{E}\{x_t|Y_t\}, \mathbf{P}_{t|t} = \mathrm{Var}\{x_t|Y_t\}$, $t = 1, 2, ..., T$. Если распределения случайных векторов модели (1)–(2) является нормальным, оценки $\mu_{t|t}$, $\mathbf{P}_{t|t}$ определяются фильтром Калмана [3].

2. Проблема сглаживания (интерполяции). Сводится к вычислению $\mu_{t-1|t}{=}\mathbf{E}\{x_{t-1}|Y_t\}$, $\mathbf{P}_{t-1|t} = \mathrm{Var}\{x_t - 1|Y_t\}$, $t = 1, \ldots, T$.

3. Проблема прогнозирования (экстраполяции). Задача прогнозирования интерпретируется как специальный случай пропущенных значений.

4. Проблема пропущенных наблюдений. Модель в пространстве состояний позволяет относительно просто решить проблему пропущенных наблюдений.

5. Проблема инициализации. Возникает, если $a_0 = \mathbf{E}\{x_0\}$ и $V_0 = \mathrm{Var}\{x_0\}$ неизвестны.

6. Оценивание параметров. В случае, когда матрицы $F_t(\theta_1)$, $H_t(\theta_2)$, $Q_t(\theta_3)$, $R_t(\theta_4)$ зависят от параметров, можно показать, что функция правдоподобия строится с помощью фильтра Калмана.

Фильтр Калмана [3] позволяет рекуррентно вычислять оценки $\mu_{t|t}{=}\mathbf{E}\{x_t|y_t\}$ и $\mathbf{P}_{t|t}{=}\mathrm{Var}\{x_t|y_t\}$. Ведем следующие обозначения:

$$x_{t|t-1} = \mathbf{E}\left\{x_t | Y_{t-1}\right\},$$
$$y_{t|t-1} = \mathbf{E}\left\{y_t | Y_{t-1}\right\},$$
$$V_{t|t} = \mathbf{E}\left\{(x_t - x_{t|t})(x_t - x_{t|t})^T\right\},$$
$$V_{t|t-1} = \mathbf{E}\left\{(x_t - x_{t|t-1})(x_t - x_{t|t-1})^T\right\},$$
$$M_{t|t-1} = \mathbf{E}\left\{(y_t - y_{t|t-1})(y_t - y_{t|t-1})^T\right\}$$
$$\vartheta_t = y_t - y_{t|t-1} = y_t - H_t x_{t|t-1}.$$

Здесь $x_{t|t-1}$ и $y_{t|t-1}$ являются прогнозными значениями x_t и y_t, полученными по наблюдениям $\{y_\tau\}$ до момента $t-1$ включительно. Матрицы $V_{t|t}$, $V_{t|t-1}$, $M_{t|t-1}$ представляют собой матрицы ковариаций ошибок оценивания; ϑ_t – остатки регрессии y_t на предыдущие наблюдения.

Используя введенные выше обозначения, удобно фильтр Калмана записать в следующей форме [4]:

$$x_{t|t} = x_{t|t-1} + K_t \vartheta_t, \tag{3}$$
$$K_t = V_{t|t-1} H_t^T \left(H_t V_{t|t-1} H_t^T + R_t\right)^{-1}, \tag{4}$$
$$V_{t|t} = (\mathbf{I} - K_t H_t) V_{t|t-1}, \tag{5}$$
$$x_{t+1|t} = F_t x_{t|t}, \tag{6}$$
$$V_{t+1|t} = F_t V_{t|t} F_t^T + Q_t, \tag{7}$$
$$M_{t|t+1} Y_{t+1} V_{t+1|t} H_{t+1}^T + R_t, \tag{8}$$
$$\vartheta_t = y_t - H_t x_{t|t-1}. \tag{9}$$

Для применения фильтра Калмана требуется задание начальных значений. Более того, чтобы применить формулы (3), (4) в момент времени $t = 1$, необходимо знать $x_{1|0}$ и $V_{1|0}$. Оценка $x_{1|0}$ означает оптимальную оценку x_1 при условии, что нет никакой информации, т. .е. наблюдения отсутствуют. Эта проблема инициализации решается просто, если даны параметры a_0 и P_0, в этом случае $x_{1|0}{=}a_0$, $V_{1|0}{=}P_0$. В случае, если эти параметры неизвестны, полагаем $x_{1|0}=0$, $V_{1|0}$ – диагональная матрица, где по диагонали ∞ (или достаточно большие числа).

II. Пропущенные значения и проблема прогнозирования

В [5] показано, что проблема пропущенных наблюдений легко решается для временных рядов, представленных в форме линейных моделей в пространстве состояний. Если наблюдение y_t пропущено в момент времени t, где $t{=}2, ..., T{-}1$, полагаем в формулах (3)–(9) $\vartheta_t = 0$, $K_t = 0$.

Проблема прогнозирования для временных рядов, представленных в форме линейных моделей в пространстве состояний, сводится к проблеме пропущенных наблюдений. Предположим, что необходимо построить прогноз для $y_{T+1}, y_{T+2}, ..., y_{T+mK}$ по наблюдениям $\{y_\tau\}$, $\tau{=}1, ..., T$, и вычислить среднеквадратические ошибки прогнозирования. Наблюдения $y_{T+1}, y_{T+2}, \ldots, y_{T+K}$ трактуются как пропущенные и используются формулы (3)–(9) для случая пропущенных наблюдений. Используем $H_{T+2} x_{T+1|T}, H_{T+2} x_{T+2|T}, \ldots, H_{T+K} x_{T+K|T+K-1}$ как прогнозные значения для $y_{T+1}, y_{T+2}, ...,$ y_{T+K} и $V_{T+1|T}, V_{T+2|T+1}, ..., V_{T+K|T+K-1}$ как матрицы ковариаций ошибок прогнозирования.

III. Прогнозирование временных рядов AR(2)

Рассмотрим AR(2)-модель временного ряда в стандартной форме

$$y_t - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} = \varepsilon_t, \tag{10}$$

где ε_t – гауссовская случайная величина с нулевым математическим ожиданием и дисперсией σ^2, φ_1, φ_2 – неизвестные параметры.

Введем обозначения

$$x_t = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix},$$

тогда модель (10) может быть представлена в форме модели в пространстве состояний

$$x_{t+1} = \begin{pmatrix} \varphi_1 & \varphi_2 \\ 1 & 0 \end{pmatrix} x_t + \begin{pmatrix} \varepsilon_{t+1} \\ 0 \end{pmatrix}, \tag{11}$$

$$y_t = \begin{pmatrix} 0 & 1 \end{pmatrix} x_t, \tag{12}$$

Модель (11)–(12) – это модель временного ряда в форме пространства состояний. где матрица $F_t{=}\begin{pmatrix} \varphi_1 & \varphi_2 \\ 1 & 0 \end{pmatrix}$ известна не полностью, а матрица $H_t{=}\begin{pmatrix} 1 & 0 \end{pmatrix}$ известна полностью, $\eta_t{=}0$. Требуется по наблюдениям $y_1, ..., y_T$ построить прогноз x_{t+K} и y_{t+K} в момент времени $t + K$, $K > 0$. Задача сводится к нахождению

$$x_{t+K|t} = \mathbf{E}\{x_{t+K} | y_t\}, \quad y_{t+K|t} = \mathbf{E}\{y_{t+K} | y_t\},$$

которые находятся по формулам (3)–(9), причем предварительно строится функция правдоподобия для нахождения оценок параметров φ_1, φ_2, σ^2. Результаты компьютерного моделирования показывают эффективность данного подхода.

1. Kalman, R. E. Contribution to the Theory of Optimal Control / R. E. Kalman // Bullet Soc. Math. Mech. – 1960. – V. 5, № 1. – P. 102–109.
2. Летов, А. М. Аналитическое конструирование регуляторов I-IV / А. М. Летов // Автоматика и телемеханика. – 1961. – № 4. – с. 436–441; № 5. – с. 526–531; № 6. – с. 661–665; 1961. – № 4. – с. 425-435.
3. Kalman, R. E. New Results in Linear Filtering and Prediction Theory / R. E. Kalman, R. Busy // J. Basic Eng. Trans. ASME. – 1961. V. 83. – P. 95–108.
4. Hamilton, J. Time Series Analysis / J. Hamilton // Princeton – 1994.
5. Durbin, J. Time Series Analysis by State Space Methods / J. Durbin, S. J. Koopman. – Oxford, 2001.