УДК 681.51

МОДЕЛЮВАННЯ ЧАСОВИХ РЯДІВ ЗА ДОПОМОГОЮ АВТОРЕГРЕСІЙНИХ МОДЕЛЕЙ

Т.І. Демківська, кандидат технічних наук, доцент
Київський національний університет технологій та дизайну
Є.О. Демківський, кандидат технічних наук, доцент
Київський національний університет імені Тараса Шевченка

Ключові слова: прогнозування поведінки системи, ряд спостережень, ARIMA-моделі, математичне сподівання, дисперсія, коваріаційна матриця.

Розглядається задача прогнозування поведінки системи, що описана часовим рядом.

Дано ряд спостережень x_1, …, x_n. Будемо вважати, що даний ряд може бути описаний в загальному випадку таким чином:

$$\nabla_r x_{t+1} = \varphi_1 \nabla_r x_{t-l_1} + \ldots + \varphi_p \nabla_r x_{t-l_p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_r \varepsilon_{t-q},$$

або

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q},$$

де t дискретний момент часу, $\varphi_1, \ldots, \varphi_p$, $\theta_1, \ldots, \theta_q$– коефіцієнти, $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$– випадковий шум з нульовим математичним сподіванням і скінченною дисперсією .

Потрібно отримати модель, яка адекватно описує вхідний ряд. Визначити структуру та оцінки параметрів цієї моделі. Будемо вважати, що даний ряд може бути описаний за допомогою ARIMA-моделі.

Запропонований нижче алгоритм призначено для розв’язання задачі оцінювання як структури, так і коефіцієнтів ARIMA-моделі. Його може бути умовно розділено на 4 частини:

- знаходження оцінок коефіцієнтів $\varphi_1, \ldots, \varphi_p$;
- знаходження оцінок коефіцієнтів $\theta_1, \ldots, \theta_q$;
- визначення початкових значень ε_{t-1}, …, ε_{t-q} для $t = t_0$;
- оцінка адекватності отриманих моделей.

Щоб уникнути ситуації переповнення при обчислювальному процесі (тобто в випадках, коли значення n доволі велике, а ряд спостережень x_i може приймати достатньо великі значення) пропонуємо нормувати ряд x_i так, щоб $x_i \in [0, 1]$. Для цього кожний елемент ряду x_i поділимо на $\max\{|x_1|, \ldots, |x_n|\}$.

Для ряду спостережень x_i значення q вибирається з припущення, що максимальний номер запізнення $l_p \le m$, також треба мати на увазі, що $n > 2m + 1$. Побудуємо коваріаційну матрицю $R_{m+1,m+1}$, де

$$r_{ij} = \sum_{k=t}^{n} x_{k-i} x_{k-j}, \; i = 0, \ldots, m; \; j = 0, \ldots, m.$$

Визначення множини структур запізнень. Виходячи з властивостей матриці R, обчислимо $V=R^{-1}$ за методом квадратного кореня .

Для визначення структури $\Phi = (\varphi_i)$ обчислюються часткові коефіцієнти кореляції $cor_i = -v_{1,i} / \sqrt{v_{11} v_{ii}}$ і відсортуємо отримані cor_i за спаданям. Для підвищення достовірності отриманих результатів пропонується ряд x_i розглядати з деяким початковим шумом, тобто $\bar{x}_i = x_i + \varepsilon_i$, де $\varepsilon_i \in [-h, h]$ рівномірно розподілена випадкова величина.

Подальший вибір параметрів шуканої моделі зводиться до розгляду різних комбінацій з множини номерів запізнень (позначимо їх як $l_1, \ldots, l_s$).

Визначення номерів коефіцієнтів запізнень адекватної моделі. Для кожної моделі обчислюємо критерій krt= $\Sigma\varepsilon_i^2 / (n - p)^{\ln(n)}$, та вибираємо моделі з найменшими значеннями krt.

Пошук оцінок $\theta_1, \ldots, \theta_n$. Виходячи з практичних міркувань далі розглядались випадки $r = 0$, $r = 1$, $r = 2$, тобто не порушуючи загальності можна вважати, що на практиці більше 3^x шумів в задачах не зустрічається. Отже, для кожної відібраної на попередньому кроці моделі досліджуємо наступні ряди:

$x_t = \varphi_1 x_{t-l1} + \varphi_2 x_{t-l2} + \ldots + \varphi_p x_{t-lp} + \varepsilon_t$;

$x_t = \varphi_1 x_{t-l1} + \varphi_2 x_{t-l2} + \ldots + \varphi_p x_{t-lp} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$;

$x_t = \varphi_1 x_{t-l1} + \varphi_2 x_{t-l2} + \ldots + \varphi_p x_{t-lp} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$;

Обчислення оцінок ε_{t_0-1}; ε_{t_0-2}. Розв’язок задачі проводиться за допомогою методу S^2.

Вибір кращої моделі за допомогою множини статистичних параметрів [1 - 2]. У цю множину входять такі статистичні параметри: t – статистика Стьюдента, яка визначає значимість кожного коефіцієнта моделі в статистичному смислі; коефіцієнт детермінації (R^2), який для адекватної моделі повинен бути близький до 1; сума квадратів помилок моделі*RSS* (для адекватної моделі приймає мінімальне значення); інформаційний критерій Акайке (*AIC*) - для адекватної моделі приймає мінімальне значення; критерій Байеса-Шварца (BSC) – анлогічний AIC, але використовується для довгих вибірок; статистика Дарбіна-Уотсона (DW) – для кращої моделі $DW = 2$;Статистика Фішера (F) – більшому значенню F відповідає більш адекватна модель, коефіцієнт Тейла (U) є важливим індикатором точності моделі. Його величина знаходиться між 0 і 1. Для U=1 модель не може бути використана для прогнозу.

При виборі найкращої моделі серед розглянутої множини моделей можна застосовувати наступний консолідований критерій адекватності, який враховує вищенаведені статистики:

$$KK = e^{1-R^2} + \frac{RSS}{N} + \left\{ \begin{matrix} \ln(AIC + BSC), & AIC + BSC > 0 \\ e^{AIC+BSC}, & AIC + BSC \le 0 \end{matrix} \right\} + e^{2-DW} + e^{U}.$$

Для кращої моделі *KK* має мінімальне значення.

Список використаних джерел

1. Бокс Дж. Анализ временных рядов: монография / Дж. Бокс, Г. Дженкинс. – М.:Мир, 1974. – 406 с. – ISBN 0-13-060774 – 6.

2. Бідюк П.І. Часові ряди: моделювання і прогнозування: монографія / П. І. Бідюк, О.І. Савенков, І.В. Баклан. – К.: ЕКМО, 2003. – 144 с.