УДК 519.6

О.П. Гожий, І.О. Калініна, В.О. Гожий

ПОБУДОВА ДИНАМІЧНИХ ПРОГНОЗІВ В ЗАДАЧАХ ПЛАНУВАННЯ

Аннотация. Проанализированы главные типы методов прогнозирования. Предложено вычисляется-тельную процедуру построения динамических прогнозов при решении задач планирования на основе анализа динамики временных рядов. Определен подход к анализа нелинейности и подход к проверке на стационарность.

Вступ. Планування реальних процесів в складних системах має бути динамічним і здатним в режимі реального часу реагувати на події, які впливають на майбутній розвиток процесу, який планується. Динамічне планування дозволяє моделювати можливі майбутні ситуації без впливу на процеси, які плануються з врахуванням невизначеностей та ризиків на кожному етапі планування і в цілому. Динамічне планування здійснюється на найвищому рівні планування в процесі визначення основних вимог до процесу, який планується, та визначення головних задач, цілей та стратегій проблеми яка вирішується. Невід'ємною частиною планування є процес прогнозування, за допомогою якого прогнозуються основні показники при побудові динамічного плану.

Постановка задачі. Метою роботи є: 1 – розглянути особливості побудови динамічних прогнозів в задачах планування; 2 – розробити алгоритм динамічного прогнозування; 3 – вирішити задачі виявлення нелінійностей та перевірку на стаціонарність.

Огляд сучасних методів і моделей прогнозування. Сучасні методи прогнозування можливо розділити на три групи [1,2,3,4]:

1. Методи прогнозування на основі розміркувань, тобто, прогнозування, що базується на суб'єктивних судженнях (експертних оцінках), інтуїції, поглиблених знаннях конкретної області та іншій інформації, що має відношення до прогнозованого процесу – так зване передбачення;

---

2. методи прогнозування на основі використання часового ряду однієї змінної, тобто з ковзним середнім (АРКС) та АРКС плюс модель тренду;

3. методи прогнозування на основі використання часових рядів декількох змінних [4,5].

В останньому випадку ендогенна змінна, що прогнозується, залежить від декількох регресорів або екзогенних змінних у правій частині рівняння. Очевидно, що в загальному випадку метод прогнозування може поєднувати у собі 2-3 наведених вище методи. На сьогоднішній день відома велика кількість методів прогнозування на основі використання часових рядів. Найбільш поширеними серед них є метод групового урахування аргументів (МГУА), методи на основі регресійного аналізу (з ковзним середнім (АРКС)), з інтегрованим ковзним середнім (АРІКС), лінійна та нелінійна множинна регресія, квантильна регресія, регресійні дерева), нейромережі, байєсівські мережі (статичні та динамічні), нечіткі множини, нечіткі нейромережі та інші.

В загальному випадку прогноз може бути представлений одним (точковим) значенням змінної, інтервалом, в який попадає випадкова змінна, а також ймовірністю прийняття змінною (чи подією) деякого значення у вибраному інтервалі. Якщо для опису процесу застосовують лінгвістичні змінні, то прогнозом буде нечітке лінгвістичне значення, але його також можна перетворити в чітке число.

Можна по-різному ставити задачу прогнозування в залежності від рівня прийняття рішення та конкретної поставленої задачі управління чи контролю. Прогнозування може стосуватись таких складових процесу:

– детермінованого тренду, як індикатора довгострокових змін процесу;

– випадкового (нерегулярного) тренду, як показника коротко- та середньострокових змін;

– короткострокових змін, тобто, прогнозування коливань (відхилень), що накладаються на тренд;

– сезонних ефектів;

– приростів (швидкості) зміни процесу, які визначаються першими різницями;

 ISSN 1562-9945

– дисперсії або стандартного відхилення, як міри розсіювання процесу (наприклад, волатильність, яку часто використовують за міру ризику у інвестуванні або міру якості на виробництві);

– якісних змінних (за допомогою нечітких множин, мереж Байєса та інш.);

– комбінацій вказаних елементів процесів.

Відповідно до того, які складові процесу необхідно прогнозувати, ставиться задача побудови математичної, ймовірнісної (ймовірнісно-статистичної) або логічної моделі, що має меті забезпечити високу якість прогнозу на заданому горизонті. Розглянемо деякі можливості математичного опису складових процесів різної природи.

Побудова алгоритму динамічного прогнозування. Для реалізації СППР при прогнозуванні динаміки часових рядів в задачах планування вибирається технологія на основі застосування методів структуризації задач, методів попередньої обробки даних, математичних і статистичних моделей процесів, множини методів оцінювання моделей і множини критеріїв визначення якості прогнозів. Застосування такого підходу забезпечує отримання високої якості прогнозів та прийнятих рішень, які на них ґрунтуються. Схема процесу аналізу даних та динамічного прогнозування в СППР на основі часових рядів представлена на рис.1.

Алгоритм процесу аналізу та динамічного прогнозування на основі часових рядів при підтримці прийняття рішень за схемою представлено в таблиці 1:

Таблиця 1

Склад алгоритму процесу аналізу та прогнозування

Кроки алгоритму	Процедура	Склад процедури
1	Попередня обробка та аналіз даних	Заповнення пропусків; згладжування екстремальних значень; логарифмування; нормування в діапазоні від –1 до +1; диференціювання (можливості обчислення перших різниць та різниць вищих порядків); пряме і зворотне перетворення Фур'є;

		цифрова фільтрація; бутстреп (розмноження вибірки даних). Перехід на крок 2.
2	Перевірка наявності нелінійно-стей	Перевірка наявності нелінійностей визначається за допомогою тесту Фішера або кореляційних функцій вищих порядків. Якщо процес містить вивода, то переходимо до кроку 3, інакше до кроку 4.
3	Визначення порядку вивода	Побудова моделі за МГВА і лінійна апроксимація процесу з метою її порівняння з нелінійною моделлю. Перехід на крок 9.
4	Перевірка процесу на стаціонарність	Перевірка процесу на стаціонарність за допомогою тесту Дікі-Фуллера. Якщо процес стаціонарний, переходимо до кроку 9, інакше до кроку 5.
5	Перевірки наявності гетероскеда-стичноті	Для перевірки наявності гетероскедастичноті застосовуються тести: Уайта, Бройша-Пагана/Годфрі, Голдфельда-Квандта. Наявність тренду визначається за допомогою тесту Дікі-Фуллера. Аналіз коінтегрованості процесів виконується за методикою Інгла-Грейнджера або Йохансена. Якщо процес гетероскедастичний, то переходимо до пункту 6, якщо містить тренд, то до кроку 7, інакше до кроку 8.
6	Визначення типу моделі	Визначення типу моделі для опису гетероскедастичності: УАРУГ(p,q), експоненційна УАРУГ(p,q), УАРУГ-М або інша. Після вибору найбільш підходящої моделі переходимо до кроку 9.
7	Визначення способу вилучення або моделювання тренду	Визначення способу вилучення або моделювання тренду включає в себе визначення порядку інтегрованості процесу або можливістю описання тренду поліноміальною функцією, експоненціальною, логарифмічно або іншими функціями. Після вилучення тренду переходимо до кроку 9.

 ISSN 1562-9945

8	Побудова моделі корекції похибок	Для коінтегрованих процесів будується модель корекції похибок і переходимо до кроку 10.
9	Побудова моделі часового ряду	Будується модель часового ряду та обчислюються критерії адекватності отриманої моделі і переходимо до кроку 10.
10	Побудова функції прогнозування	На основі обраної моделі, будується функція прогнозування. Обчислюється прогноз поведінки ряду та визначаються оцінки точності прогнозу.

Визначення наявності нелінійностей. Для розв'язання цієї задачі можна користуватися різними критеріями. Однак при цьому необхідно знати про їх можливості. Покажемо на простому прикладі, що застосування лінійних коваріаційних функцій не завжди призводить до позитивних результатів [1,7,10].

Нехай при визначенні структури моделі не були обчислені деякі пояснюючі змінні; у результаті залишки описуються таким рівнянням:

$$\xi(k) = c\,u(k-1)\,e(k-1) + e(k)\,, \tag{1}$$

де $e(k)$ – білий гаусівський шум; $E[e(k)] = 0,\ E[u(k)] = 0,\ E[e(k)u(k)] = 0$, тобто, змінні $e(k)$ і $u(k)$ некорельовані і мають нульове середнє; c – масштабний коефіцієнт. Можна показати, що нормована автокореляційна функція залишків із нормованої функції взаємної кореляції між вхідним сигналом $u(k)$ і залишками мають вигляд: $\Phi_{\xi\xi}(\tau) = \delta(\tau), \quad \Phi_{u\xi}(\tau) = 0, \quad \forall\tau$.

Проте, із рівняння (1) випливає, що $\xi(k)$ – це послідовність, що буде вносити зміщення в оцінки параметрів моделі. Таким чином, у загальному випадку лінійні кореляційні методи не дають можливості встановити факт наявності нелінійних ефектів та ступінь їх впливу на процес.

Для того щоб оцінити тип зв'язку між входом і виходом (тобто, зв'язок лінійна або нелінійна) можна скористатися спектральною функцією високого порядку вигляду:

$$X_{ij} = \frac{\left|S_{\omega}(\omega_i, \omega_j)\right|^2}{S_{\omega}(\omega_i)S_{\omega}(\omega_j)S_{\omega}(\omega_i / \omega_j)},$$

де $S_{\omega}(\omega_i, \omega_j)$ – біспектральна щільність потужності; $S_{\omega}(\omega_i)$ – спектральна щільність потужності часового ряду. При $S_{\omega}(\omega_i, \omega_j) = 0, \ \forall \omega_i, \omega_j$ процес буде лінійним і третій момент вхідного сигналу $\mu_3 = 0$. Проте, якщо $X_{ij} = const$, то процес лінійний, але $\mu_3 \neq 0$.

Такий підхід до встановлення наявності нелінійностей має два недоліки. По-перше, оцінювання спектральної щільності потужності потребує застосування спеціального попередньої обробки сигналів у вигляді застосування часових вікон, усереднення, цифрової фільтрації і т.п. По-друге, він не завжди може бути використаний при розв’язанні задач ідентифікації систем, оскільки він не дає можливості одержати оцінки параметрів моделі в явному вигляді. Крім того, при розв’язанні цих же задач не завжди є можливість одержати виміри вхідного сигналу або ж інформативний вхідний сигнал одержують штучно у вигляді спеціально генерованих послідовностей, що не завжди можна подавати на вхід об’єкта внаслідок особливостей його функціонування.

Що стосується економічних процесів, то в цьому випадку, як правило, не можна поставити експеримент із процесом. Тому використовують тільки ті статистичні дані, які можна реально зібрати в процесі дослідження. У загальному випадку при ідентифікації систем використовують три типи сигналів: вхідні, вихідні і збурення. При цьому вхідний керуючий сигнал вважають незалежним від збурення. У результаті виявляється неможливим з’ясувати деякі типи зв’язків.

Можливо використання також дисперсійного методу визначення присутності нелінійностей, який заснований на застосуванні наступної функції:

$$\Psi_{zu}(t_1, t_2) = E_{u(t_2)}[E_{z(t_1)}[z(t_1) \mid u(t_2)] - E_{z(t_1)}[z(t_1)]]^2,$$

яка обчислюється за допомогою достатньо складаного інтегрального рівняння, якщо відомі відповідні щільності розподілу можливостей сигналів, що не завжди можна визначити.

 ISSN 1562-9945

Рисунок 1 - Схема процесу динамічного прогнозування на основі часових рядів

Крім розглянутих підходів до визначення наявності нелінійностей, при побудові регресійних моделей можна скористатися простішими тестами. Наприклад, статистикою Фішера:

$$\widehat{F} = \frac{\dfrac{1}{k-2}\sum_{i=1}^{k}\ \sum_{j=1}^{n_i} n_i\,(\overline{y}_i - \widehat{y}_{ij})^2}{\dfrac{1}{n-k}\sum_{i=1}^{k}\ \sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2},$$

де k – число груп даних; n_i – кількість вимірів у групі; $\overline{y}_i$ – групове середнє; $\widehat{y}_i$ – значення, що оцінюють по прямій регресії; n – загальне число вимірів. Фактично дана статистика являє собою таке відношення:

$$\widehat{F} = \frac{\text{Відхилення середніх значень від прямої регресії}}{\text{Відхилення значень } y(k) \text{ від групових середніх}}.$$

Якщо статистика $\widehat{F}$ зі $\nu_1 = k-2$, $\nu_2 = n-k$ ступенями свободи досягає або перевершує рівень значущості, то гіпотезу про лінійність потрібно відкинути.

Перевірка процесу на стаціонарність за допомогою тесту Дікі – Фуллера. При визначенні наявності вивода (тобто присутності одиничного кореня), пропонується скористатися тестом Дікі-Фуллера [6], суть якого полягає в наступному: для визначення наявності одиничного кореня запропонвано скористатись трьома рівняннями:

$$\Delta y(k) = \gamma\, y(k-1) + \varepsilon(k), \tag{2}$$

$$\Delta y(k) = a_0 + \gamma\, y(k-1) + \varepsilon(k), \tag{3}$$

$$\Delta y(k) = a_0 + \gamma\, y(k-1) + a_2 k + \varepsilon(k), \tag{4}$$

де k – дискретний час; $\gamma = a_1 - 1$ – коефіцієнт у рівнянні:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k).$$

Різниця між рівняннями (2 та (3), (4) полягає у присутності детермінованих членів a_0 і $a_2 k$ у рівняннях (3) і (4), відповідно. Рівняння (2) представляє собою модель випадкового кроку (або "блукання"), друге включає зміщення у вигляді константи a_0, а третє включає зміщення та детермінований лінійний часовий тренд.

У всіх трьох рівняннях нас цікавить параметр γ. Якщо $\gamma = 0$, то послідовність $\{y(k)\}$ містить одиничний корінь. Застосування тесту Дікі-Фуллера передбачає оцінювання одного або більше з наведених вище трьох рівнянь за допомогою МНК або ММП з метою отримання оцінки параметра γ та стандартної похибки цієї оцінки. На основі оцінки та її стандартної похибки обчислюється t –статистика, яка порівнюється із значеннями, наведеними в таблицях Дікі-Фуллера. На основі цього порівняння приймається рішення щодо справедливості або відхилення нуль-гіпотези, що $\gamma = 0$.

Нехай для рівняння $y(k) = a_1 y(k-1) + \varepsilon(k)$ на основі 100 спостережень отримана така оцінка параметра $\widehat{a}_1 = 0,9459$ із стандартною похибкою $SE_{a_1} = 0,031$. Очевидно, що оцінювання рівняння

 ISSN 1562-9945

$\Delta y(k) = \gamma\, y(k-1) + \varepsilon(k)$ приведе до оцінки $\hat{\gamma} = -0,0541$ з тією ж стандартною похибкою 0,031. Таким чином, для нуль-гіпотези $\gamma_0 = 0$ відповідна t –статистика буде дорівнювати: $t = -0,0541 / 0,031 = -1,74516$.

Із таблиць Дікі-Фуллера для випадку $a_0 = a_2 = 0$ при $N = 100$ знаходимо, що критичні значення t –статистики дорівнюють −1,61; −1,95 та −2,60 на рівнях значущості 10%, 5% та 1%, відповідно. Таким чином, в розглянутому гіпотетичному випадку при $\hat{\gamma} = -0,0541$ нуль-гіпотеза щодо $\gamma = 0$ (тобто, одиничний корінь присутній) не може бути відхилена при рівнях значимості 5% та 1%, але вона відхиляється на рівні значущості 10%. Як і у більшості інших випадків перевірки гіпотез, для будь-якого рівня значущості, критичні значення t –статистики зменшуються при збільшенні розміру вибірки.

Методика тестування на наявність одиничного кореня залишається незмінною для всіх трьох рівнянь (7.29)-(7.31). Однак, критичні значення t –статистики залежать від структури моделі, тобто від того чи наявні у моделі зміщення α_0 та детермінований тренд $\alpha_2 k$. Автори методики визначили, що критичні значення для $\gamma = 0$ залежать від структури регресійного рівняння та від довжини вибірки. Так, для рівняння (2) використовується статистика, що позначається через τ, для рівняння (3) – статистика позначається τ_μ, а для рівняння (4) – через τ_τ.

Якщо у модель включити константу (зміщення), але $a_2 = 0$, то необхідно користуватись іншою частиною таблиці критичних значень для t –статистики. Оцінюючи рівняння для гіпотетичного прикладу у формі: $\Delta y(k) = a_0 + \gamma\, y(k-1) + \varepsilon(k)$, знайдемо, що $\gamma = 0,9135 - 1 = -0,0865$ із стандартною похибкою 0,041. Таким чином, отримаємо таке значення t –статистики: $\tau_\mu = -0,0865 / 0,041 = -2,1098$.

З таблиць знову знаходимо, що для 100 спостережень критичні значення дорівнюють −2,58; −2,89 та −3,51 на рівнях значущості 10%, 5% та 1%, відповідно. Таким чином, нуль-гіпотеза щодо наявності одиничного кореня ($\gamma = 0$) не може бути відхилена при всіх рівнях значущості, закладених в таблицю Дікі-Фуллера.

Якщо ж скористатись структурою моделі у вигляді: $\Delta y(k) = a_0 + a_1 y(k-1) + a_2 k + \varepsilon(k)$, то знайдемо, що критичні значення статистики τ_τ дорівнюють −3,45 та − 4,04 на рівнях значущості 5% і 1%, відповідно.

Критичні значення не зміняться, якщо рівняння (2), (3) і (4) замінити такими рівняннями вивода:

$$\Delta y(k) = \gamma\, y(k-1) + \sum_{i=2}^{p} \beta_i\, \Delta y(k-i+1) + \varepsilon(k), \tag{5}$$

$$\Delta y(k) = a_0 + \gamma\, y(k-1) + \sum_{i=2}^{p} \beta_i\, \Delta y(k-i+1) + \varepsilon(k), \tag{6}$$

$$\Delta y(k) = a_0 + \gamma\, y(k-1) + a_2 k + \sum_{i=2}^{p} \beta_i\, \Delta y(k-i+1) + \varepsilon(k). \tag{7}$$

Тобто при використанні цих рівнянь для тестування нуль-гіпотези $\gamma = 0$ використовують ті ж статистики: τ, τ_μ і τ_τ. Для перевірки об'єднаних гіпотез щодо коефіцієнтів Дікі і Фулер запропонували ще три $F-$ статистики: φ_1, φ_2 і φ_3.[8,9]

Так, статистика φ_1 використовується для перевірки нуль-гіпотези щодо $\gamma = a_0 = 0$ в рівняннях (3) і (6). Статистика φ_2 використовується для перевірки об'єднаної гіпотези $a_0 = \gamma = a_2 = 0$ в рівняннях (4) і (7), а статистика φ_3 − для перевірки об'єднаної гіпотези $\gamma = a_2 = 0$.

Статистики φ_1, φ_2 і φ_3 обчислюються по аналогії із звичайними $F-$ статистиками:

$$\varphi_i = \frac{[RSS_1 - RSS_2]\, r}{RSS_2 (N - n)},$$

де RSS_1 і RSS_2 − суми квадратів похибок (СКП), обчислених для моделей з обмеженнями та моделей без обмежень; $r-$ число обмежень; $N-$ число використаних спостережень; $n-$ число параметрів, оцінених для необмеженої моделі. Можливими обмеженнями можуть бути обмеження на порядок моделі та їх структуру.

ISSN 1562-9945

Обчислення значення φ_i та його порівняння із відповідним значенням, запропонованим Дікі і Фулером, дозволяє визначити рівень значущості, на якому обмеження на модель відіграють суттєву роль. При цьому за нуль-гіпотезу приймають те, що дані генеруються моделлю з обмеженнями, а за альтернативну – що дані генеруються моделлю без обмежень. Якщо обмеження не відіграє суттєвої ролі, то сума квадратів похибок для моделі з обмеженнями буде близькою до СКП для моделі без обмежень. Відповідно, φ_i буде мати при цьому невелике значення, тобто, при великих значеннях φ_i обмеження відіграють суттєву роль і нуль-гіпотеза відхиляється.

Таким чином, якщо розраховане значення φ_i є меншим ніж відповідна статистика Дікі і Фуллера, то приймається модель з обмеженнями (тобто, приймається нуль-гіпотеза). Якщо ж розраховане значення φ_i є більшим ніж відповідна статистика Дікі і Фуллера, то нуль-гіпотеза відхиляється і приймається альтернативна, тобто, що обмеження грають суттєву роль і модель повинна бути без обмежень.

Необхідно також виконати перевірку гіпотез щодо значущості константи a_0 та коефіцієнта a_2, значущість якого означає наявність часового тренду. Якщо нуль-гіпотеза визначена як $\gamma = 0$, то тестування на наявність часового тренду виконується за допомогою статистики $\tau_{\beta\tau}$. Таким чином, ця статистика виконує перевірку чи $a_2 = 0$ при $\gamma = 0$. Для перевірки гіпотези стосовно $a_0 = 0$, необхідно скористатись статистикою $\tau_{\alpha\tau}$, якщо оцінюється модель (7), або статистикою $\tau_{\alpha\mu}$, якщо оцінюється модель (6).

Висновки. Представлено алгоритм прогнозування на основі аналізу динаміки часових рядів. Визначено підхід до аналіза нелінійностей та підхід до перевірки на стаціонарність. Алгоритм призначено для вирішення задач побудови динамічних прогнозів. У подальших дослідженнях необхідно удосконалити структуру та продовжити аналіз моделей прогнозування, які є перспективними для оцінювання та прогнозування в задачах динамічного планування і визначитись з найкращими критеріями якості оцінок прогнозів.

ЛИТЕРАТУРА

1. Бідюк П.І. Моделювання та прогнозування нелінійних динамічних процесів / Бідюк П.І., Баклан І.В., Баклан Я.І., Коршевнюк Л.О. та інш. – К.:ЕКМО, 2004. – 120 с.
2. Бідюк П.І. Системний підхід до побудови математичних моделей на основі часових рядів / Бідюк П.І., Баклан І.В., Рифа В.М. // Системні дослідження та інформаційні технології, №3, 2002. – с. 114-131.
3. Бідюк П.І. Часові ряди: моделювання та прогнозування / Бідюк П.І., Савенков О.І. Баклан І.В. – Київ: ЕКМО, 2004. – 144 с.
4. Бідюк П.І. Системний підхід до прогнозування на основі моделей часових рядів // Системні дослідження та інформаційні технології, 2003, № 3, с. 88-110.
5. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. – М.: Финансы и статистика, 2003. – 414 с.
6. Chatfield C. Time series forecasting.– London: Chapman & Hall, 2000 –267 p.
7. Zgurovs- ky M.Z., Bidyuk P.I., Terentyev O.M. Method of constructing Bayesian networks based on scoring functions // Cybernetics and System Analysis, 2008, Vol. 44, No.2, pp. 219-224.
8. http://www.mataf.net/en/tools/home
9. Altman E.I., Avery R.B., Eisenbeis R.A., Sinkey J. Application of Classification Techniques in Business, Banking and Finance. – Greenwich: JAI Press, 1981. – 418 p.
10. Hosmer D.W., Lemeshow S. Applied Logistic Regression. – New York: John Wiley & Sons, Inc., 2000. – 380 p.

 ISSN 1562-9945