

## МАТЕМАТИЧНИЙ АПАРАТ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ ЕКОНОМІЧНОГО ТА ЕКОЛОГІЧНОГО ТИПІВ, ЩО МОЖУТЬ БУТИ ПІДДАНІ ЗОВНІШНІМ ВПЛИВАМ

**Постановка проблеми.** Проблема прогнозування часових рядів була і залишається актуальною, особливо останнім часом, коли стали доступними потужні засоби для збору і обробки інформації. Разом з цим підвищилися вимоги до процесу прогнозування і ускладнилися залежності часових рядів.

Проблема прогнозування відноситься до слабо структурованих проблем, тому до кожної задачі потрібно підходити окремо, що зумовлює велику різноманітність методів прогнозування. Але число базових методів прогнозування, які в тих чи інших варіаціях повторюються в інших методах, набагато менше. Багато з цих «методів» належать до окремих прийомів або процедур прогнозування, інші відрізняються від базових або один від одного кількістю приватних прийомів і послідовністю їх застосування.

Різноманітні дані в техніці, економіці, соціології, медицині та інших сферах надходять у вигляді часових рядів. Вони є, як правило, нестационарними, оскільки їх основні характеристики змінюються в часі. Основою для прогнозування служить історична інформація, що зберігається в інформаційних сховищах у виді часових рядів. Якщо можна побудувати математичну модель і знайти шаблони, що адекватно відбивають цю динаміку, є імовірність, що з їх допомогою можна передбачати і поведінку системи в майбутньому. Прогнозування часових послідовностей дозволяє на основі аналізу поведінки часових рядів оцінити майбутні значення прогнозованих змінних.

Прогнозування на основі часових рядів – один із самих популярних підходів до прогнозування розвитку економічних процесів, об'ємів торгових операцій, об'ємів виробництва та накопичення продукції на складах, оцінювання альтернативних економічних стратегій, формування бюджетів підприємств та держави, прогнозування та менеджмент економічних і фінансових ризиків та інше. Загалом методи прогнозування можна розділити на два широкі класи: 1) методи прогнозування на основі використання часового ряду однієї змінної, тобто, на основі авторегресії, авторегресії з ковзним середнім (АРКС) та АРКС плюс модель тренду; 2) методи прогнозування на основі використання часових рядів декількох змінних.

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.

**Аналіз публікацій за темою дослідження.** Але майже всі відомі методи прогнозування часових рядів будь-якої природи базуються на внутрішній природі поведінки процесу. Виходять з внутрішніх закономірностей. Ці методи не дозволяють враховувати так звані «зовнішні впливи», що можуть призводити до змін поведінки процесу (зміну характеристик ряду, тренду і таке ін.). Тобто за своїми якісними характеристиками поведінку таких часових рядів можна цілком вести на засадах феноменологічних підходів [1]. До такого типу часових рядів можна віднести ряди економічних показників, або ряди спостережень за екологічними процесами. Для економічних рядів, це, наприклад, курси цінних паперів, або валют, ціни на банківські метали, індекси ділової активності. Для часових рядів такого роду зовнішніми впливами можуть бути інсайдерська інформація, що може вплинути на курси цінних паперів, або заяви державних діячів, що можуть привести до різкого стрибка курсу місцевої валюти відносно резервних валют. Для екологічних процесів, це можуть бути аномальні явища, що погано описуються за допомогою динаміко-аналітичних підходів, а через відсутність сталих періодів протікання, стає неможливим використання статистичних підходів. Але їхні прояви можуть суттєво впливати на навколишнє середовище. Це такі природні явища як цунамі, землетруси, виверження вулканів, зсуви та ін. Тож постає необхідність визначення математичного апарату, що можна застосовувати для прогнозування такого типу часових рядів навіть при наявності зовнішніх впливів.

Для прогнозування поведінки часових рядів, що є чутливими до зовнішніх впливів, пропонується застосовувати методи прогнозування часових рядів, що в змозі враховувати ці впливи. Такими математичними засобами є методи прихованих марківських моделей [2], подібних траєкторій [3], лінгвістичного моделювання [4]. Але більш точним методом прогнозування представляється комбінований або гібридний метод: лінгвістичне моделювання – приховані марківські моделі [5].

**Мета статті.** Стоїть задача визначення щодо математичного апарату для прогнозування часових рядів економічного та екологічного типів, що можуть бути піддані зовнішнім впливам.

**Приховані марківські моделі.** Діаграма, подана нижче (рис. 1), показує загальну структуру ПММ. Овали - це змінні з випадковим значенням. Випадкова змінна  $x(t)$  відповідає значенню прихованої змінної в момент часу  $t$ . Випадкова змінна  $y(t)$  - це значення змінної, за якою ми спостерігаємо, в момент

часу  $t$ . Стрілки на діаграмі символізують умовні залежності. Із діаграми можна дізнатись, що значення прихованої змінної  $x(t)$  (в момент часу  $t$ ) залежить тільки від значення прихованої змінної  $x(t-1)$  (в момент  $t-1$ ). Це називається властивістю Маркова. Хоча в той же час значення змінної  $y(t)$ , за якою ми спостерігаємо, залежить лише від значення прихованої змінної  $x(t)$  (в момент часу  $t$ ).

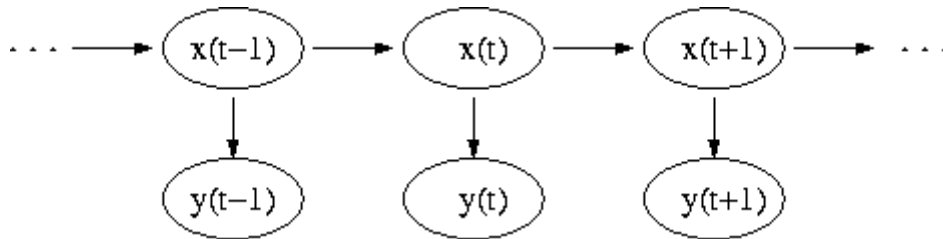


Рис. 1. Загальна схема функціонування прихованої Марковської моделі

Ймовірність спостерігати послідовність  $Y = y(0), y(1), \dots, y(L-1)$  довжини  $L$  дорівнює:

$$P(Y) = \sum_X P(Y | X) P(X)$$

тут сума пробігає по всіх можливих послідовностях прихованих вузлів  $X = x(0), x(1), \dots, x(L-1)$

Основні приховані марковські моделі можна описати за допомогою таких змінних:

- $N$  - кількість станів
- $T$  - кількість спостережень
- $\theta_{i=1, \dots, N}$  - параметр для спостереження за зв'язками між станами
- $\Theta_{i=1, \dots, N, j=1, \dots, N}$  - ймовірність переходу із стану  $i$  до стану  $j$
- $\Theta_{i=1, \dots, N}$  -  $N$ -мірний вектор, що складається із  $\Theta_{i=1, \dots, N}$
- $x_{t=1, \dots, T}$  - стан спостереження за час  $t$
- $y_{t=1, \dots, T}$  - результат спостереження за час  $t$
- $F(y | \theta)$  - функція розподілу ймовірності спостережень, параметризованих по  $\theta$ .

**Метод подібних траєкторій.** Ідея методу полягає в наступному. Маємо ряд спостережень екологічного процесу, що їх зроблено за якийсь час  $\{y(1); y(2); \dots; y(n)\}$ , графік якого наведено на рис.2.

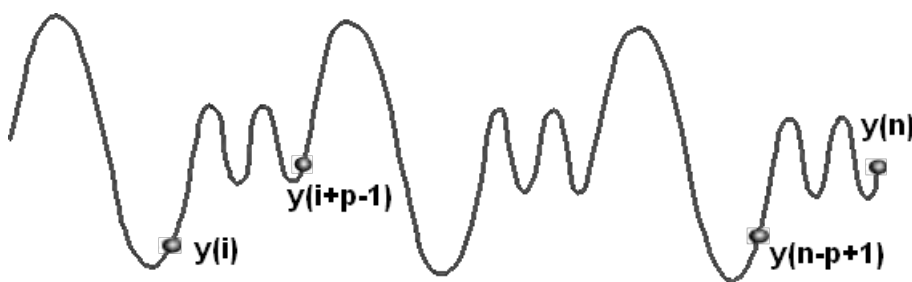


Рис.2. Динаміка ряду спостережень

Змінна  $y(i)$ ,  $i = \overline{1, N}$  тут представлена фізичними значеннями відповідного процесу (наприклад, сила вітру, інтенсивність стоку води, сила підземних поштовхів).

За обраним критерієм обирається ділянка траєкторії “найближча” до ділянки, яка передуює прогнозованій точці. Надалі оцінюється прогноз за формулою  $y(n+1) = y(i+p)$ , де

$$I = \min \left\{ \sum_{i=1}^p |y(j+i-1) - y(n-p+i)| \right\} \quad J = 1, 2, \dots, n-p;$$

$$J = \min_i |y(i+j-1) - y(n)| \quad i = I, I+1, \dots, I+p-1.$$

Формалізувати метод можна наступним чином. Нехай ми маємо наступні вектори спостережень

$$Y_1 = (y_1, y_2, \dots, y_p)^T, Y_2 = (y_2, y_3, \dots, y_{p+1})^T, \dots, Y_K = (y_k, y_{k+1}, \dots, y_{k+p+1})^T, \dots, Y_N = (y_{n-p+1}, y_{n-p+2}, \dots, y_n)^T$$

Знаходимо найближчу точку із умови мінімальної відстані

$$Y_k = \arg \min_j d(Y_n, Y_j).$$

Є й інші способи пошуку найближчої точки, наприклад, найбільш поширена метрика – квадрат евклідової відстані

$$d(Y_k, Y_n) = (Y_k - Y_n)^T (Y_k - Y_n).$$

### **Лінгвістичне моделювання**

За певними правилами числові значення послідовності замінюються символами [7]. Прогнозування відбувається шляхом пошуку ланцюжків символів і співставлення їх з ланцюжками з бази даних спостережень. Отримані символи утворюють алфавіт, з якого, в свою чергу, формуються слова. Таким чином маємо:

$\Sigma$  : - множина термінальних символів

$N$  : - множина нетермінальних символів

$V = \Sigma \cup N$  : словник граматики (множина всіх термінальних та нетермінальних символів).

При цьому під граматику (або формальною граматику) розуміється спосіб опису формальної мови, тобто виділення деякого підмножини з безлічі всіх слів деякого кінцевого алфавіту.

Для прогнозування використовують стохастичну контекстно-вільну граматику [8]. Стохастична контекстно-вільна граматики (СКС) - це контекстно-вільна граматики, в якій кожному правилу виводу відповідає ймовірність.

Контекстно-вільна граматики  $G$  п'ятиціпка  $(N, T, P, S)$  :

- $S \in N$
- $N$  та  $T$  скінченні множини, що не перетинаються
- $P$  скінченна підмножина  $N \times (N \cup T)^*$

При цьому, використовують такі назви:  $N$  — множина нетермінальних символів,  $T$  — множина термінальних символів,  $P$  — множина правил виводу  $S$  початковий символ. Правила  $(\alpha, \beta) \in P$  записують як  $\alpha \rightarrow \beta$ .

В лівій частині правила виводу має знаходитись одна змінна (нетермінальний символ). Формально, має виконуватись  $\alpha \in N, \beta \in (N \cup T)^*, |\beta| \geq 1$ .

В стохастичних контекстно-вільних граматиках правилам виведення співставляють ймовірність використання:  $\rho : P \rightarrow \mathbb{R}$  де  $\sum_{p_r \in P_r} \rho(p_r) = 1$

Стохастичні контекстно вільні граматики розширюють контекстно вільні граматики так само, як приховані марківські моделі розширюють регулярні граматики.

**Гібридний метод: лінгвістичне моделювання – приховані марковські моделі.** Основними принципами гібридизації є: принцип неоднорідності, принцип плюралізму, принцип системного аналізу неоднорідного завдання, принцип конструктора, принцип пріоритету знань, принцип поступовості, перший та другий принцип спадкування, принцип самоорганізації агрегованої моделі, принцип повноти зниження продуктивності агрегованої моделі. Ці принципи певним чином узагальнюють світову практику досліджень розробки гібридних інтелектуальних систем прийняття рішень. Використання цих принципів дозволить робити агреговані моделі там де це доцільно, такими які вони мають бути, та функціонуючими згідно з очікуванням.

Класичні гібридні системи комбінують аналогові та дискретні моделі, Агрегативні системи моделюють аналітико-статистичні закономірності біхаверістики складних систем. Методологія інтелектуальних експертних систем дозволяє перебороти недоліки символічного підходу за рахунок комбінування із традиційними інформаційними та технологіями штучного інтелекту. У кожному із цих випадків мова йде про становлення нової інформаційної методології моделювання складних процесів і явищ шляхом побудови і застосування моделей-гібридів. Зазначимо що існують різні шляхи для досягнення поставленої мети.

Використання наведеної гібридної ЛМ-ПММ моделі дозволяє враховувати зовнішні впливи, що можуть призводити до змін характеристик часового ряду, його тренду тощо.

**Заключення.** До недоліків наведеного математичного апарату можна віднести загальний недолік всіх методів, де використовуються статистичні дані – брак історичної інформації. В такому разі прогнозування відбувається на основі суджень, тобто, прогнозування, що ґрунтується на суб'єктивних судженнях (оцінках), інтуїції, поглиблених знаннях конкретної області та іншій інформації, що має відношення до прогнозованого процесу – це так зване передбачення. В разі відсутності достатньої кількості таких даних, необхідно використовувати методи системного аналізу [6]. Методи системного аналізу є найбільш універсальні і адекватні сучасним вимогам засоби дослідження. У свою чергу це дає можливість особі, що приймає рішення, одержати максимально повну, достовірну і головне, своєчасну інформацію про можливі зовнішні впливи. Використання системного аналізу також дає можливість враховувати зовнішні впливи на поведінку відповідних процесів..

**Висновки та перспективи подальших досліджень.** Математичний апарат для прогнозування поведінки часових рядів вже достатньо розвинутий. Але наявність зовнішніх впливів вимагає застосування таких підходів, що дають можливість враховувати невидимі чинники, які вносять своє збурення. Визначено напрямок для подальших досліджень запропонованих методів прогнозування часових рядів, що можуть піддаватися зовнішнім впливам.

**ЛИТЕРАТУРА:**

1. Гуссерль Э. Логические исследования.[Текст] Т. 2. Исследования по феноменологии и теории познания./ Гуссерль Э. — М.: ДИК, 2001.. — 470 с.
2. Баклан І.В. Класифікація моделей Марковського типу [Текст] Наукова монографія/ Баклан І.В., Степанкова Г. А. – К.: НАУ, 2012. - 84 с.
3. Баклан І.В. Структурний підхід до розпізнавання образів у системах безпеки / І.В. Баклан, Ю.М. Селін, О.О. Петренко //Національна Безпека України: стан, кризові явища та шляхи їх подолання. Міжнародна науково-практична конференція (Київ, 7-8 грудня 2005 р.). Збірка наукових праць. – К.: Національна академіяуправління – Центр перспективних соціальних досліджень. 2005. – С.375-380.
4. Баклан І.В. Лінгвістичне моделювання: основи, методи, деякі прикладні аспекти. [Текст] / І.В. Баклан // Системные технологи. Региональный межвузовский сборник научных трудов. – Выпуск 3(74) – 2011, с.10-19.
5. Баклан І.В. Імовірнісні моделі для аналізу та прогнозування часових рядів [Текст] / І.В. Баклан, Г.А. Степанкова // Штучний інтелект. — 2008. — № 3. – С.505-515.
6. Згуровский М. З., Панкратова Н. Д. Системный анализ. Проблемы. Методология. Применение. – Київ. – Науково-виробниче підприємство «Наукова думка» НАН України, 2011.- 726 с.
7. Селін Ю.М. Системне оцінювання і прогнозування аномальних природних процесів марківського типу [Текст] : дис. кандидата техн.. наук : 01.05.04 / Селін Юрій Миколайович. – К., 2009. – 176 с.
8. Clark, A. (2001). "Unsupervised Language Acquisition: Theory and Practice" (PhD thesis) [Електроннийресурс].Режим доступу: <http://arxiv.org/abs/cs/0212024v1>

БАКЛАН Ігор Всеволодович – к.т.н., доцент, завідувач кафедри інтелектуальних систем Національної академії управління.

Наукові інтереси:

- лінгвістичне моделювання, приховані марковські моделі.

СЕЛІН Юрій Миколайович – к.т.н., доцент кафедри інтелектуальних систем Національної академії управління.

Наукові інтереси:

- лінгвістичне моделювання, приховані марковські моделі.