

УДК 004:519.2

О.Ю. Берзлев

ДВНЗ «Ужгородський національний університет», Ужгород

СУЧАСНИЙ СТАН ІНФОРМАЦІЙНИХ СИСТЕМ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Сформульовано основні постановки задачі прогнозування часових рядів. Розглянуто переваги і недоліки, а також окреслено перспективи розвитку класичних статистичних та інтелектуальних прогнозних моделей та методів, які складають основу сучасних інформаційних систем прогнозування часових рядів.

Ключові слова: модель прогнозування, часовий ряд, інформаційна система

Сформулированы основные постановки задачи прогнозирования временных рядов. Рассмотрены преимущества и недостатки, а также очерчены перспективы развития классических статистических и интеллектуальных прогнозных моделей и методов, которые составляют основу современных информационных систем прогнозирования временных рядов.

Ключевые слова: модель прогнозирования, временной ряд, информационная система

Basic problem statements of time series forecasting are formulated. Advantages and disadvantages are considered and, in addition, the prospects of the development of classic statistic and intellectual forecast models and methods which form the basis of current information systems of time series forecasting are delineated.

Keywords: the model of forecasting, time series, information system

Постановка проблеми

Прогнозування – одна з найбільш складних і в той же час найбільш актуальних і затребуваних задач аналізу даних (Data Mining). Складність процесу прогнозування пов'язана з необхідністю аналізу і оцінювання великих обсягів даних, ускладненням методів, появою концептуально нових підходів до прогнозування процесів різної природи тощо. Тому на сьогодні стан розвитку методів прогнозування тісно пов'язаний з розвитком інформаційних технологій. Так звані, інформаційні системи прогнозування, що відображають цей зв'язок в рамках економетрики, фінансової математики, статистики, набувають свого прояву в широкому спектрі прикладних галузей науки, а також у сферах виробництва, фінансового планування в економіці і торгівлі. Сьогодні вони є невід'ємними складовими процесів управління складними системами і прийняття управлінських рішень, застосовуються аналітиками для оцінювання ризиків фінансового інвестування тощо.

Під інформаційною системою прогнозування будемо розуміти взаємопов'язану сукупність програм, кожна з яких виконує певні функції: отримання вхідної інформації та її обробка, формування вихідних даних на основі визначених алгоритмів, метою яких є

реалізація прогнозу відповідно до цілей особи, яка приймає рішення.

Основними складовими системи прогнозування часових рядів є база з ретроспективними даними, яка безперервно поповнюється, комплекс прогнозних моделей, а також методів оцінювання їх якості, що згруповані залежно від постановки задачі прогнозування. Функціонування такої системи здійснюється в діалоговому режимі з особою, яка приймає рішення.

Якість будь-якої інформаційної системи прогнозування визначається ефективним управлінням процесом прогнозування. Під процесом прогнозування будемо розуміти аналіз і оцінювання, на основі визначених наукових підходів тенденцій розвитку певного процесу або явища, використовуючи наявну інформацію про перебіг даного процесу або явища в минулому, тобто ретроспективні дані. Перебіг розвитку процесів, зокрема соціально-економічних, може фіксуватися у визначені моменти часу. Як правило, фіксуються тільки найбільш значущі характеристики, які визначають перехід процесу з одного стану в інший. Таким чином, більшість процесів може представлятися за допомогою часових рядів, тобто послідовністю значень деяких величин, які вимірюються у визначені моменти часу. Будь-який часовий ряд визначається двома обов'язковими

складовими: значенням ряду або вимірюванням та моментом часу, в який це значення тим або іншим способом було зафіксоване.

Розробка ефективних інформаційних систем прогнозування часових рядів є актуальною задачею як для теорії, так і для практики в різних галузях. Зокрема в галузі економіки і фінансів необхідність їх застосування можна пояснити високим ступенем мінливості розвитку економічних систем, який відбувається в умовах невизначеності, нестійкості та ризику, передумовою чого є багато факторів, серед яких: тенденції глобалізації, ускладнення господарських взаємозв'язків, ріст темпів розвитку національних ринків тощо. Нестабільність економічних процесів ускладнює застосування класичних статистичних прогнозних методів, які часто складають основу систем прогнозування, через їх малу ефективність в цих умовах. Це обумовлює необхідність розробки нових і модифікації відомих методів та моделей прогнозування тенденцій і показників розвитку економічних процесів, побудови методів для підвищення ефективності аналізу, оцінювання і моделювання ризику, впровадження інформаційних технологій, які б своєчасно реагували на зміни в динаміці розвитку економічних систем, економили час, характеризувалися б високою точністю та гнучкістю.

Актуальність цих досліджень має не тільки практичне значення для прогнозистів, інвесторів та аналітиків, але й теоретичне: розробка концепцій і математичного інструментарію, що може скласти основу для подальших наукових досліджень в цій галузі.

Аналіз останніх досліджень і публікацій

Аналіз сучасних робіт виявив, що попри наявність великої кількості публікацій з даної тематики, тільки невелика частина із запропонованих прогнозних моделей і відповідних методів можуть бути ефективно застосовані на практиці як складові систем прогнозування. Окрім класичних статистичних моделей [1-10], останнім часом для підвищення точності прогнозування часових рядів досліджуються комбіновані багаторівневі моделі [6;11;12] та методи інтелектуального аналізу [13-15], зокрема кластеризації, індексації, класифікації, ідентифікації викидів тощо.

Мета статті

Метою статті є огляд і характеристика відомих прогнозних моделей, які можуть складати основу інформаційних систем прогнозування, а також формулювання пропозицій щодо підвищення ефективності управління процесом прогнозування.

Формальні постановки задачі прогнозування часових рядів

Сформулюємо означення часового ряду. Нехай S – дискретна множина. Дискретним часовим рядом

$$\{z_i\}_{i=1}^n = \{z_1, z_2, \dots, z_n\} = \{z(t_1), z(t_2), \dots, z(t_n)\}$$

будемо називати скінченну послідовність вимірювань, які проводяться в дискретні моменти часу $t_i \in S$, $i = \overline{1, n}$, $n \in \mathbb{N}$ і фіксують певні характеристики зміни станів процесу або явища, яке досліджується; t_1 – початковий момент часу. Будемо вважати, що вимірювання проводяться через певні часові інтервали: хвилина, година, день, тиждень, місяць, квартал тощо. Результати вимірювань дійсні $z_i \in \mathbb{R}$, $i = \overline{1, n}$, $n \in \mathbb{N}$ і можуть являти собою однорідні економічні показники (рівень інфляції, об'єм випущеної продукції), фінансові показники тощо. В економетриці часто розглядають часовий ряд як часткову реалізацію стохастичного процесу, який представляється нескінченною випадковою послідовністю. Але на відміну від елементів випадкової вибірки, значення часового ряду не є незалежними та однаково розподіленими.

Можна виділити такі основні задачі прогнозування часових рядів

- прогнозування майбутніх значень;
- прогнозування знаків приростів часових рядів;
- ідентифікація моментів локальних екстремумів.

Перша задача формулюється так: на основі ретроспективних значень $z_n, z_{n-1}, \dots, z_{n-m+1}$, $m \leq n$

ряду $\{z_i\}_{i=1}^n$, найбільш точно оцінити його поведінку

в майбутньому в моменти $t_{n+1}, t_{n+2}, \dots, t_{n+\theta}$, тобто побудувати послідовність прогнозних значень

$$\{\hat{z}_i\}_{i=n+1}^{n+\theta} = \{\hat{z}_{n+1}, \hat{z}_{n+2}, \dots, \hat{z}_{n+\theta}\},$$

де θ – горизонт прогнозування; m – об'єм ретроспективної вибірки.

Позначимо через $\hat{z}_\tau(n)$ прогноз, який розраховується в момент t_n (в точці n) на τ точок вперед, $\tau = \overline{1, \theta}$. Така функціональна залежність, яка дозволяє описати поведінку часового ряду, називається моделлю прогнозування. Позначимо модель для даної задачі через F_1 . Прогноз часового

ряду $\{z_i\}_{i=1}^n$ на одну точку вперед на основі даної моделі можна формально записати у вигляді

$$\hat{z}_{n+1} = \hat{z}_1(n) = F_1(z_{n-m+1}, z_{n-m}, \dots, z_n).$$

У випадку

прогнозування з горизонтом $\theta > 1$, застосовують ітераційний підхід:

$$\begin{aligned}\hat{z}_{n+1} &= \hat{z}_1(n) = F_1(z_{n-m+1}, z_{n-m+2}, \dots, z_n), \\ \hat{z}_{n+2} &= \hat{z}_2(n) = F_1(z_{n-m+2}, z_{n-m+3}, \dots, z_n, \hat{z}_{n+1}), \\ &\vdots \\ \hat{z}_{n+\theta} &= \hat{z}_\theta(n) = F_1(z_{n-m+\theta}, \dots, \hat{z}_{n+\theta-2}, \hat{z}_{n+\theta-1}), \quad m \leq n.\end{aligned}$$

Залежно від типу моделі, вона може мати ряд параметрів, які потребують окремої оцінки. До таких моделей належать, наприклад, адаптивні поліноміальні моделі [4; 6]. Побудувати найбільш точну оцінку означає побудувати таку модель, яка задовольняє відповідній критерій оцінки якості прогнозування. Для даної задачі такими критеріями можуть бути середнє абсолютне відхилення, середня квадратична похибка, стандартне відхилення, відносна похибка тощо.

На фінансовому і валютному ринках, часто досить передбачати короткочасну динаміку часового ряду без розрахунку безпосередньо прогнозних значень. Проблемою є побудова такої найточнішої моделі, яка б визначала прогноз знаку приросту значення часового ряду на одну точку вперед. Дана задача, зазвичай, застосовується для визначення напрямку руху ціни валютних пар і може використовуватися для визначення екстремальних точок або точок розвороту ринку, тобто таких точок, які вказують подальший напрямок руху ціни. Модель такого типу може розглядатися як частковий випадок попередньої задачі, проте на практиці для підвищення точності прогнозування знаків приростів застосовують специфічні моделі, наприклад, модель Лукашина [6].

Позначимо моделі цього типу через F_2 .

Розглянемо формальну постановку задачі прогнозування знаків приростів часового ряду. На

основі ряду $\{z_i\}_{i=1}^n$, побудуємо ряд, який складається

з перших різниць $\{\Delta z_i\}_{i=2}^n$,

де $\Delta z_i = z_i - z_{i-1}$, $i = \overline{2, n}$.

Позначимо через $\{\chi_i\}_{i=2}^n$ знаковий ряд, кожен елемент якого задається формулою: $\chi_i = \text{sgn}(\Delta z_i)$ або

$$\chi_i = \begin{cases} +1, & \Delta z_i > 0 \\ 0, & \Delta z_i = 0 \\ -1, & \Delta z_i < 0 \end{cases} \text{ при } i = \overline{2, n},$$

а прогноз знаку приросту, який розраховується в точці n на τ точок вперед позначимо через $\hat{\chi}_\tau(n)$.

Тоді модель F_2 у випадку прогнозування на одну точку вперед визначається так:

$$\hat{\chi}_{n+1} = \hat{\chi}_1(n) = F_2(z_{n-m+1}, z_{n-m+2}, \dots, z_n).$$

Має місце також й ітераційний підхід у випадку прогнозування знаку приросту для $\theta > 1$, проте на практиці його не застосовують. В даній задачі, на відміну від попередньої, використовуються специфічні критерії оцінки якості, які часто будуються на основі функції Хевісайда або шляхом підрахунку кількості правильно передбачених знаків приростів з врахуванням значущості розрахованих прогнозів за допомогою вагових коефіцієнтів.

Ще однією задачею, яка використовується на фінансових ринках є ідентифікація моментів локальних екстремумів [12]. Для цієї задачі будують спеціальні методи, які часто засновані на математичних інструментах технічного аналізу, серед яких трендові індикатори плинних середніх, індикатори сили [16] тощо.

Слід зазначити, що окрім наведених задач, розглядають постановки з врахуванням зовнішніх факторів, які можуть здійснювати вплив на явище або процес, що досліджується.

Огляд моделей прогнозування, їх переваги та недоліки

Дослідження статей, опублікованих у 15 провідних наукових виданнях у 70-80-тих рр., проведене Тафтом [17] показало, що понад 75% графічної інформації представляється часовими рядами. Завдяки появі потужних процесорів стало можливим не тільки представляти перебіг процесів за допомогою часових рядів, але й швидко та з необхідною точністю прогнозувати їх розвиток в майбутньому. Впродовж останніх тридцяти років теорія прогнозування інтенсивно розвивається. Розроблено щонайменше сотня різних моделей, ще більше методів, і з кожним роком їх кількість зростає. Проте на практиці зазвичай застосовують не більше 30 основних моделей, інші застосовні у вузькоспеціалізованих задачах або є непопулярними через низьку точність. Основними класичними моделями прогнозування типу F_1 , які застосовуються впродовж багатьох років в широкому спектрі задач як менеджерами для прийняття управлінських рішень, так і дослідниками в різних галузях є:

- моделі експоненціального згладжування, які через наочність і простоту часто застосовують аналітики для короткострокового прогнозування. До них належать [1-4]: адаптивна поліноміальна модель Брауна, модель Хольта, яка застосовується для моделювання часових рядів з вираженою трендовою складовою, модель Вінтерса, яка застосовується для рядів з адитивною трендовою складовою та мультиплікативною сезонністю, адаптивна модель прогнозування Трігга-Ліча, в якій значення параметра згладжування динамічно коригується на кожному кроці залежно від похибки прогнозування на основі плинного контрольного сигналу (tracking signal).

Теоретичні і емпіричні аспекти вибору оптимальних параметрів згладжування моделей наведені в роботах [5; 6];

- регресійні моделі, які призначені для прогнозування або оцінювання значення однієї залежної змінної на основі значень інших незалежних змінних шляхом встановлення між ними зв'язку. Розрізняють парну регресію, в якій залежна змінна оцінюється на основі єдиної незалежної та множинна, в якій розглядають комплекс незалежних змінних. Також виокремлюють лінійну та нелінійну регресії;

- авторегресійні моделі, ідея яких полягає у ідентифікації залежності між значеннями або спостереженнями часового ряду шляхом дослідження автокореляції між спостереженнями, які вибирають з певним фіксованим часовим інтервалом [4]. Класичними моделями, які належать до даного класу є авторегресійна модель порядку p AR (p), авторегресійна модель плинного середнього порядків p та q ARMA (p, q) [4]. Вважається, що для застосування наведених моделей, вхідний ряд має бути стаціонарним, тобто мати сталі математичне сподівання і дисперсію. Розвитком моделі ARMA (p, q) стали авторегресійні моделі інтегрованого плинного середнього (ARIMA (p, d, q)) і ARIMAX, які були запропоновані в роботі [3] і використовуються у випадку, коли часовий ряд являє собою нестационарний процес. Подальшим розвитком авторегресійних моделей стали спеціалізовані моделі умовної гетероскедастичності, які застосовуються передусім на фінансовому ринку, в тому числі для аналізу волатильності часових рядів: ARCH, основою якої є твердження про залежність умовної дисперсії від квадратів попередніх значень часового ряду та GARCH (узагальнена ARCH), згідно з якою умовні дисперсії залежать не тільки від попередніх значень ряду, але й від попередніх значень умовної дисперсії [7]. У 90-тих рр. XX ст. було розроблено багато модифікацій моделі GARCH: GARCH-M, EGARCH, AGARCH, QGARCH і т.д.

Слід зазначити, що кожна з наведених моделей і відповідні їм методи мають як переваги, так і недоліки. Наприклад, використання моделей експоненціального згладжування обумовлене їх простотою і прозорістю, в той же час накладені на кожну з моделей вимоги обмежують їх широке застосування. Регресійні моделі теж характеризуються простотою і гнучкістю [8], проте і нелінійна, і лінійна регресійні моделі мають свої недоліки. Лінійна регресійна модель не використовується для моделювання нелінійних процесів, а як відомо більшість часових рядів, для яких виникає задача прогнозування і аналізу, характеризуються нелінійністю і нестійкістю відносно середнього рівня. У той же час, складність ідентифікації функціональних залежностей та розрахунку параметрів нелінійних регресійних

моделей в багатьох випадках обмежує їх застосування [9].

Останнім часом, порівняно часто застосовують авторегресійні моделі та методи прогнозування для багатьох прикладних задач. Модель ARIMA або модель Бокса-Дженкінса, яка є однією з базових моделей даного класу, давно стала класичною. Для неї розраховані оцінки параметрів залежно від особливих властивостей часових рядів, вона гнучка та легко моделюється [3]. Проте суттєвим недоліком цієї моделі є складність розрахунку численних параметрів, її обмеженість вимогами стаціонарності часових рядів, нормальності та незалежності залишків [3; 10] і, як наслідок, нездатність моделювати нелінійні процеси. Наведені вимоги не будуть виконуватися для більшості прикладних часових рядів, зокрема рядів, які виникають на фінансовому ринку. Механізм генерації часових рядів не завжди є лінійним і стаціонарним.

Слід зазначити, що залежно від задачі прогнозування, точність моделі може мати кілька аспектів, зокрема похибка прогнозування безпосередньо значень часових рядів на визначений горизонт, похибка прогнозування знаків приростів. Крім того, у випадку прогнозування фінансових показників, точність може розглядатися і через призму подальшого прийняття рішень. В цьому випадку модель або метод, що їй відповідає, буде характеризуватися з точки зору досягнутої економічної ефективності. Саме тому, ефективні моделі або методи прогнозування повинні володіти здатністю ідентифікувати ключові характеристики часових рядів і використовувати їх для досягнення поставлених цілей. Вказані класичні моделі цієї здатності не мають.

Можна виділити два основні напрямки досліджень, які останнім часом сформувалися в галузі створення систем прогнозування:

- створення комбінованих багаторівневих моделей прогнозування;
- розробка та використання методів інтелектуального аналізу часових рядів.

Підходи до конструювання комбінованих моделей і їх експериментальні дослідження наводяться в роботах [6; 11; 12]. Ефективні методи адаптивного оцінювання моделей прогнозування, які складають базу комбінованих моделей, дозволяють в цілому підвищити точність прогнозування, проте їх побудова пов'язана зі значними труднощами. Побудований за комплексом прогноз на основі селективного або гібридного підходів, повинен бути точнішим, ніж прогноз окремо взятої моделі. Точність в цьому випадку досягається завдяки безперервному оцінюванню багатьох адаптивних параметрів.

Іншим напрямком, який користується популярністю, є інтелектуальний аналіз часових рядів або Time-Series Data Mining, під яким розуміють набір інноваційних методів, який включає підходи інтелектуального аналізу даних, що призначені для

ідентифікації прихованих або асоціативних правил даних часових рядів, методи нелінійної динаміки [18], оптимізації на основі генетичних алгоритмів та програмування генетичних виразів [19] і т.д. З моменту виникнення інтелектуального аналізу часових рядів розроблено більше сотні спеціальних методів, які стосуються задач індексації, кластеризації, класифікації [13; 14], ідентифікації викидів тощо.

Висновки і перспективи подальших досліджень

Проблеми створення систем прогнозування в першу чергу стосуються побудови такого комплексу моделей, який би характеризувався необхідною точністю, простотою, гнучкістю в застосуванні і прозорістю в оцінці параметрів. Застосування більшості класичних статистичних методів і моделей пов'язане з виконанням ряду жорстких вимог: стаціонарності, нормальності та незалежності залишків часових рядів. Конструювання нелінійних або комбінованих моделей пов'язане зі значною складністю і не завжди приводить до підвищення точності. Використання методів інтелектуального аналізу теж характеризується певними недоліками, зокрема в задачах кластеризації виявляється неоднозначність результатів при виборі різних мір близькості. Тим не менше, вказані моделі та методи за умови ефективного управління процесом прогнозування, зокрема поєднання в комплексах статистичних моделей та інтелектуальних методів, можуть бути успішно застосовані та адаптовані до цілей прогнозних інформаційних систем.

На погляд автора, перспективним напрямом в розробці інформаційних систем прогнозування на сьогодні є побудова комбінованих адаптивних моделей, які можуть включати адаптивні поліноміальні моделі різних порядків, моделі авторегресії тощо, переваги яких видобуваються з результатів кластеризації часового ряду, зокрема методами найближчого сусіда або К-найближчих сусідів.

Слід зазначити, що останніми роками з'являється розуміння того, що для багатьох задач інтелектуального аналізу даних необхідно застосовувати не швидкий та наблизений, а повільний, але точний і ґрунтовний пошук [15], пропонується використовувати методи, які можуть бути максимально легко практично застосовані. З огляду на це, розроблена інформаційна система повинна бути достатньо простою у застосуванні та долати обмеження і недоліки класичних підходів аналізу, зокрема тих, які стосуються стаціонарності часових рядів.

Список літератури

1. Brown Robert G. *Statistical forecasting for inventory control*. – US: McGraw-Hill Inc., 1959. – 223 p.
2. Holt Charles C. *Forecasting trends and seasonal by exponentially weighted averages // International Journal of Forecasting*. – 1957. – Vol.20, no.1.-P.5-10.
3. Box G.E.P., Jenkins G.M., *Time series analysis: forecasting and control*. – San Francisco: Holden-Day, 1976. – 575 p.
4. Vercellis Carlo. *Business intelligence: data mining and optimization for decision making*. – John Wiley & Sons, Ltd., Publication, 2009. – 417 p.
5. Gardner E. *Exponential Smoothing: the state of the art // Journal of Forecasting*. – 1985. – Vol.4. – P. 1-38.
6. Лукашин Ю.П. *Адаптивные методы краткосрочного прогнозирования временных рядов: Учеб. пособие*. – М.: Финансы и статистика, 2003. – 416 с.
7. Bollerslev Tim. *Generalized autoregressive conditional heteroskedasticity // Journal of Econometrics*. – 1986. – Vol. 31. – P. 307-327.
8. Draper N., Smith H. *Applied regression analysis*. – New York: Wiley, In press, 1981. – 693 p.
9. Тихонов Э.Е. *Методы прогнозирования в условиях рынка: учеб. пособие*. – Невинномысск, 2006. – 221 с.
10. Pandit S.M., Wu, S.-M. *Time series and system analysis with applications*. – New York: Wiley, 1983. – 586 p.
11. Берзлев О.Ю., Маляр М.М., Ніколенко В.В. *Адаптивні комбіновані моделі прогнозування біржових показників // Вісник Черкаського держ. технолог. ун-ту. Серія: технічні науки*. – 2011. – № 1. – С. 50-54.
12. Берзлев А.Ю. *Оценка эффективности прогнозирования и принятия решений на финансовом рынке // «Problems of Computer Intellectualization», V.M. Glushkov Institute of Cybernetics of NAS of Ukraine*. – Kyiv-Sofia: ITHEA, 2012. – С. 249-257.
13. Keogh E., Pazzani M. *An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback // 4th Int'l Conference on Knowledge Discovery and Data Mining*. 1998 Aug 27-31. – New York. – P. 239-241.
14. Singh S. *Pattern Modeling in Time-Series Forecasting // Cybernetics and Systems. An International Journal*. – 2000. – Vol. 31, no. 1. – P. 49-65.
15. Chang C.L.E., Garcia-Molina H., Wiederhold G. *Clustering for approximate similarity search in high-dimensional spaces // IEEE Transactions on Knowledge and Data Engineering*. 2002 Jul - Aug. – Vol 14, no.4. – P. 792 -808.
16. Стивен Б. Акелис. *Технический анализ от А до Я*. – М.: Диаграмма, 1999. – 315 с.
17. Tufte E. *The visual display of quantitative information/ 2nd edition*. – Graphics Press, 2001. – 200 p.
18. Peters E. E. *Fractal market analysis: applying chaos theory to investment and economics*. – John Wiley & Sons, Inc, 1994. – 336 p.
19. Снитюк В.С. *Прогнозування. Моделі. Методи. Алгоритми: Навчальний посібник*. К.: «Маклаут», 2008. – 364 с.

Стаття надійшла до редколегії 28.02.2013

Рецензент: д-р техн. наук, проф. Й.І. Головач, ДВНЗ «Ужгородський національний університет», Ужгород.