

Аналіз методів, моделей та програмних засобів прогнозування часових рядів

Дніпропетровський національний університет ім. Олеся Гончара

Проведено аналіз існуючих у цей час методів і моделей прогнозування часових рядів і надано їхню класифікацію. Наведено опис як класичних статистичних методів і моделей прогнозування таких, як моделі експоненціального згладжування, ARIMA моделі, регресійні моделі, так і більш сучасних, які ґрунтуються на принципі машинного навчання, наприклад, модель на основі опорних векторів (SVM), нейронні мережі, марковські моделі та класифікаційно-регресійні дерева. Описано програмні пакети, які містять модулі для аналізу та прогнозування часових рядів. Подано напрямки майбутніх досліджень і можливих програмних розробок у сфері прогнозування фінансових часових рядів.

Ключові слова: класифікація методів і моделей прогнозування часових рядів, адаптивні моделі, ARIMA моделі, регресійні моделі, SVM, марковські моделі, нейронні мережі, класифікаційно-регресійні дерева.

Вступ

Сьогодні задача прогнозування часових рядів набуває особливої актуальності в різних областях людської діяльності: у природничих науках – для прогнозування кількості опадів, стану забруднення водних ресурсів, деяких біологічних і біохімічних показників; в економіці – для прогнозування щоденних коливань цін на акції, курсів валют, щотижневих і щомісячних обсягів продажів, річних обсягів виробництва тощо.

Протягом останніх 60 років було проведено велику кількість досліджень щодо можливості прогнозування фінансових часових рядів. Ці роботи мають на меті визначити, активи яких компаній будуть підніматися у вартості та в який бізнес було б вигідно інвестувати, або, як говорять, спеціалісти, «перемогти на ринку». Це означає мати таку прибутковість, яка постійно перевищує середню прибутковість на ринку при збереженні такого ж рівня ризику, як і решта ринку. Протягом довгого періоду часу для цього використовували фундаментальний аналіз. Фундаментальний аналіз – це вивчення основних чинників, які сприяють добробуту економіки в цілому, промислових секторів та окремих компаній. Але такий підхід має певні недоліки. По-перше, такі дослідження потребують великих затрат часу, по-друге, такий аналіз обов'язково супроводжується необ'єктивністю аналітика. Тому в останній час багато спеціалістів в області фінансів використовують методи аналізу даних, зокрема прогнозування часових рядів, для розв'язання цієї задачі.

Постановка задачі

Задача прогнозування часового ряду зводиться до того, щоб, маючи певну кількість спостережень за деяким явищем $u_1, u_2, \dots, u_N$, побудувати прогноз щодо майбутньої поведінки цього процесу, тобто значення $u_{N+1}, u_{N+2}, \dots, u_{N+k}$.

Великий інтерес до прогнозування з боку спеціалістів у різних сферах знань спричинив стрімкий розвиток цієї області досліджень, що викликало появу нових методів прогнозування часових рядів та модернізацію вже існуючих. Сьогодні кількість методів і моделей прогнозування стала дуже великою. У зв'язку з цим виникає питання єдиної класифікації методів і моделей прогнозування. Крім аналізу та класифікації існуючих методів і моделей прогнозування часових рядів, у

статті буде проведено огляд програмних пакетів прогнозування часових рядів. Особливу увагу планується приділити так званим структурним моделям прогнозування, по-перше, тому, що ця область досліджень є дуже перспективною сьогодні, по-друге, тому, що ці методи можуть бути застосовані для рядів, які характеризуються нелінійністю та суттєвими коливаннями, таких як фондові ринки та фінансові часові ряди.

Аналіз літературних досліджень та основний матеріал

У дисертаційній роботі к. т. н. І.О.Чучуєвої [1] запропоновано розмежовувати поняття методу та моделі прогнозування. Під методом прогнозування автор пропонує розуміти послідовність дій, які потрібно зробити для отримання моделі прогнозування, а під поняттям моделі прогнозування – функціональне уявлення, що адекватно описує досліджуваний процес і є основою для отримання його майбутніх значень. У цій же роботі [1] наведено класифікацію методів і моделей прогнозування часових рядів. Згідно з цією класифікацією методи прогнозування часових рядів поділяють на дві великі групи: інтуїтивні та формалізовані.

Інтуїтивні методи прогнозування стосуються суджень і оцінок експертів. Сьогодні їх часто застосовують у маркетингу, економіці та політиці, тому що система, поведінка якої необхідно спрогнозувати, або дуже складна і не піддається математичному опису, або дуже проста і такого опису не потребує[2].

Формалізовані методи – це методи прогнозування, які будують моделі прогнозування часового ряду, тобто визначають таку математичну залежність між рівнями часового ряду та часом, яка дозволяє обчислити майбутнє значення процесу, тобто зробити прогноз [2].

Одночасно автор у роботі [1] пропонує класифікацію і самих моделей прогнозування часових рядів на дві групи: моделі предметної області та моделі часових рядів.

Моделі предметної області – це такі математичні моделі прогнозування, для побудови яких використовують закони предметної області. Наприклад, математична модель, яка робить прогноз щодо погоди, містить рівняння динаміки рідин і термодинаміки. Прогноз рівня цукру крові людини, хворої на діабет, здійснюють на основі системи диференціальних рівнянь. Отже, в таких моделях використовують функціональні залежності, властиві конкретній предметній області.

Моделі часових рядів – це математичні моделі прогнозування, які мають знайти залежність майбутнього значення від минулого всередині самого процесу і на цій залежності обчислити прогноз. Ці моделі є універсальними для різних предметних областей, тобто їх загальний вигляд не змінюється залежно від природи часового ряду. Наприклад, нейронні мережі можна використовувати для прогнозування температури повітря, а після цього побудовану модель на нейронних мережах застосувати для прогнозу біржових індексів.

Окрім цього, моделі часових рядів можна поділити на статистичні та структурні. У статистичних моделях залежність майбутнього значення від минулого задається у вигляді деякого рівняння [3]. До них відносяться регресійні моделі, зокрема МГУА, авторегресійні моделі, моделі на основі експоненціального згладжування та інші.

Загальна схема класифікація методів і моделей прогнозування часових рядів може бути подана у такому вигляді (рис. 1):

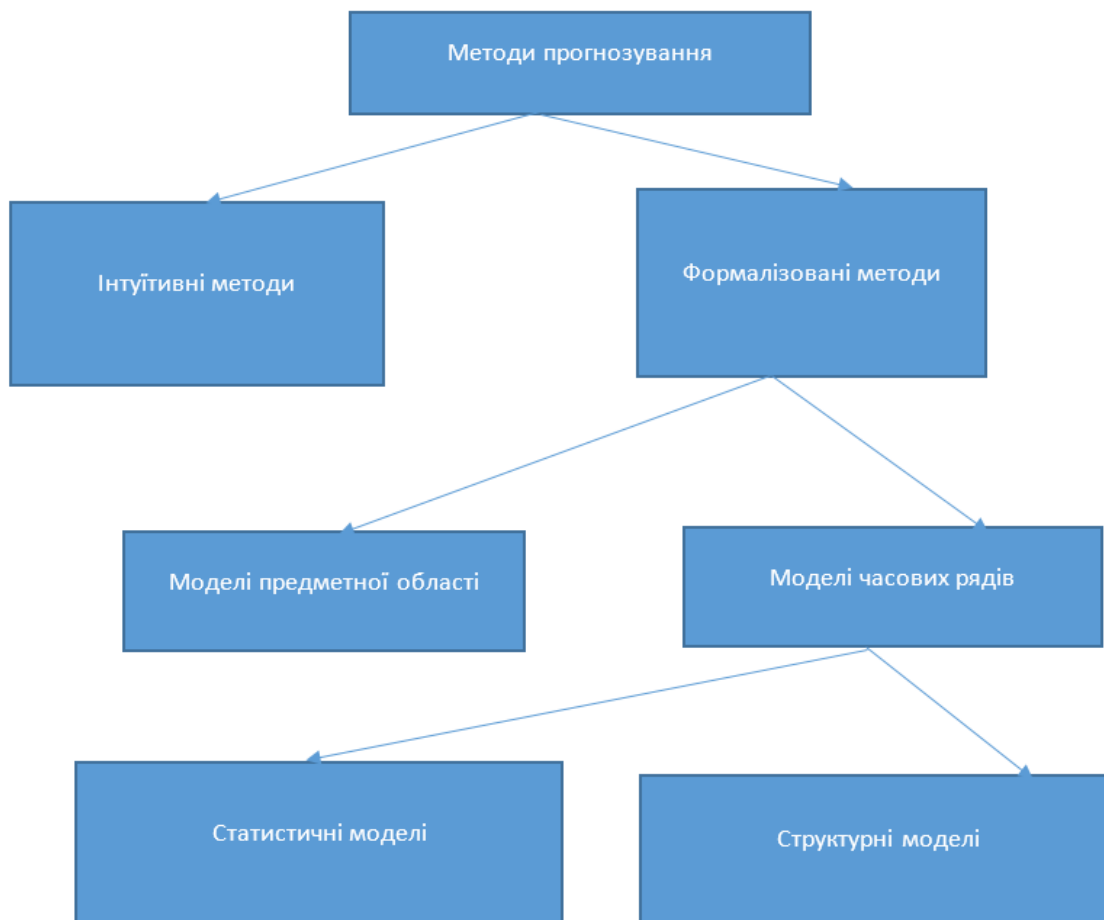


Рис. 1. Класифікація методів і моделей прогнозування часових рядів

У разі регресійного підходу значення часового ряду u_t подають у вигляді функції $u_t = F(t, a_1, \dots, a_k)$, такої, що

$$F(t, a_1, \dots, a_k) : |u_t - F(t, a_1, \dots, a_k)| \rightarrow \min_{a_1, \dots, a_k}. \quad (1)$$

Функцію $F(t, a_1, \dots, a_k)$ знаходять за допомогою лінійної та нелінійної регресії [5]. Перевагами цих методів є простота та прозорість моделювання, однаковість процесу аналізу та проектування, недоліками – складність визначення функціональної залежності; трудомісткість знаходження коефіцієнтів залежності; накладання обмежень на довжину ряду. Особливою проблемою у випадку нелінійної регресії є складність процесу ідентифікації форми кривої.

Моделі та методи прогнозування часових рядів на основі експоненціального згладжування також називають адаптивними. У цих моделях часовий ряд подають у вигляді функції

$$u_t = f(a_{1t}, a_{2t}, \dots, a_{pt}, t) + e_t, \quad (2)$$

під час побудови якої відслідковують величину відхилень прогнозних значень від значень вихідного ряду [4]. Вони передбачають виконання двох етапів: згладжування ряду та прогнозування.

Адаптивні моделі поділяють на лінійні та сезонні. Лінійні моделі враховують

лише трендову складову. В них значення рівнів ряду обчислюють у вигляді суми :

$$u_{t+\tau} = a_{1,t} + a_{2,t}\tau, \quad (3)$$

де τ – кількість кроків прогнозу; $a_{1,t}$, $a_{2,t}$ – коефіцієнти адаптивної моделі в момент часу t ; $a_{1,t}$ – оцінка того, чого досягли на поточному кроці; $a_{2,t}$ – приріст на поточному кроці.

До адаптивних моделей лінійного зростання відносять модель Хольта, модель Тейла – Вейджа, модель Брауна та модель Бокса – Дженкінса. Моделі відрізняються засобами знаходження параметрів $a_{1,t}$, $a_{2,t}$.

Адаптивні сезонні моделі дозволяють, крім тренду, враховувати й сезонну складову, яку подають у формі масивів довжиною l $g_1, \dots, g_l$, $f_1, \dots, f_l$, де l – період сезонності. Їх поділяють на чотири групи: модель лінійного зростання з мультиплікативною сезонністю, також відома як модель Уінтерса; експоненціальне зростання з мультиплікативною сезонністю; лінійне зростання з адитивною сезонністю; експоненціальне зростання з адитивною сезонністю.

Перевагами цих методів є простота моделювання; однаковість аналізу і проектування; простота математичних обчислень; відсутність жорстких обмежень на довжину тренувальної послідовності. До недоліків відносять сильну залежність якості прогнозу від параметрів згладжування, у зв'язку з чим виникає нова проблема – проблема пошуку оптимальних значень цих параметрів; неможливість урахування істотної частини історії змін значень ряду.

У середині 90-х років минулого століття був розроблений принципово новий і досить потужний клас алгоритмів для прогнозування часових рядів. Більшу частину роботи з дослідження методології та перевірки моделей було проведено двома статистиками, – Г.Е.П. Боксом і Г.М. Дженкінсом. Вони запропонували цілу ієрархію алгоритмів, найвідомішим з яких є алгоритм ARIMA. Він вбудований практично в будь-який спеціалізований пакет для прогнозування.

Модель ARIMA (p, q, d) для нестационарного часового ряду має вигляд

$$\Delta^d X_t = c + \varepsilon_t + \sum_{i=1}^p \alpha_i \Delta^d X_{t-i} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \quad (4)$$

де ε_t – стаціонарний часовий ряд; c , α_i , b_j – параметри моделі; Δ^d – оператор різниці часового ряду порядку d (послідовне взяття d раз різниць першого порядку – спочатку від часового ряду, потім від отриманих різниць першого порядку, потім від другого порядку і т.д.).

Підхід ARIMA до часових рядів полягає в тому, що в першу чергу оцінюється стаціонарність ряду. Різними тестами виявляються наявність одиничних коренів і порядок інтегрованості часового ряду (зазвичай обмежуються першим або другим порядком). Далі за необхідності (якщо порядок інтегрованості більше нуля) ряд перетворюється взяттям різниці відповідного порядку і вже для перетвореної моделі будується деяка ARMA-модель, оскільки передбачається, що отриманий процес є стаціонарним, на відміну від вихідного нестационарного процесу. У методології ARIMA задається лише загальний клас моделей, що описують часовий ряд і дозволяють якось виражати поточне значення змінної через її попередні значення. Потім алгоритм, підлаштовуючи внутрішні параметри, сам вибирає найвідповіднішу модель прогнозування.

У цей час існує декілька відомих модифікацій моделей ARIMA, такі, як SARIMA, які дозволяють урахувати сезонну компоненту; ARMAX – дає можливість враховувати вплив зовнішніх чинників; ARCH – модель з умовною гетероскедастичністю; GARCH – узагальнена модель ARCH; AGARCH – асиметрична модель GARCH; TGARCH – порогова модель GARCH; IGARCH – інтегрована модель GARCH та ін. [5]

Перевагами моделей прогнозування ARIMA є простота та прозорість моделювання; однаковість аналізу і проектування та різноманітність сфер застосування; недоліками – трудомісткість і ресурсомісткість ідентифікації найбільш відповідної моделі та неможливість моделювання нелінійних залежностей.

У структурних моделях залежність майбутнього значення від минулого задається у вигляді певної структури і правил переходу за нею. До них відносяться нейромережні моделі, моделі на базі ланцюгів Маркова, моделі на базі класифікаційно-регресійних дерев та ін.

Багато досліджень в області прогнозування часових рядів присвячено питанню побудови нейронних мереж для цього. Штучні нейронні мережі являють собою систему з'єднаних і взаємодіючих простих процесорів (штучних нейронів). Кожен процесор подібної мережі має справу тільки із сигналами, які він періодично отримує, і сигналами, які він періодично посилає іншим процесорам. Проте з'єднані в досить велику мережу, такі локально прості процесори разом здатні виконувати досить складні завдання [6].

Нейрон являє собою одиницю оброблення інформації в нейронній мережі. На рис. 2 нижче показано модель нейрона, що лежить в основі штучних нейронних мереж.

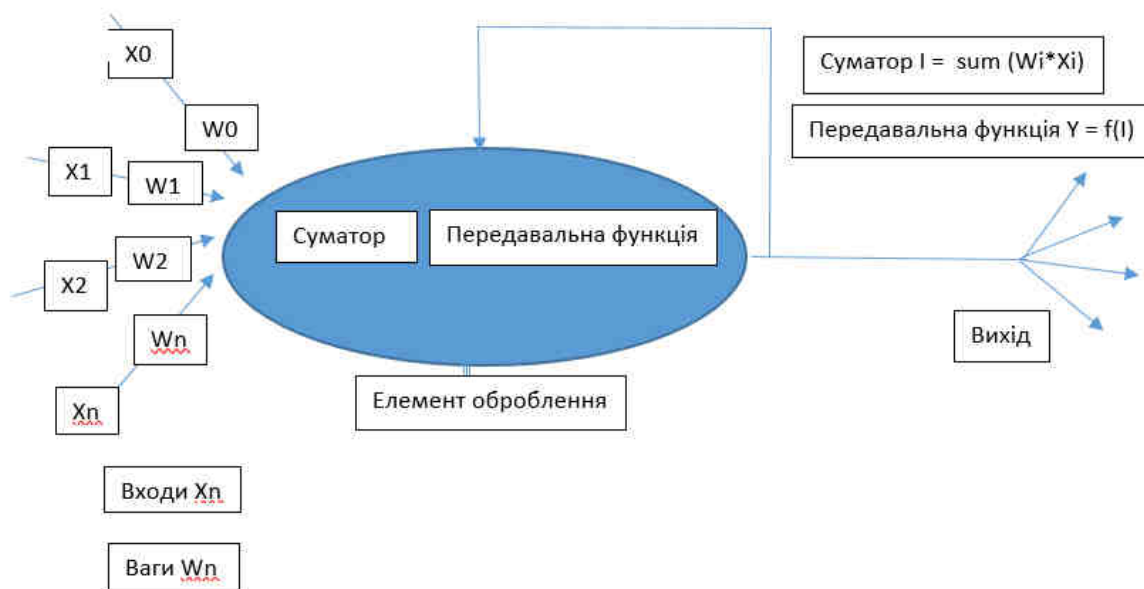


Рис.2 Модель нейрона

Математично нейрон являє собою ваговий суматор, єдиний вихід якого визначається через його входи і матрицю ваг таким чином:

$$y = f(u),$$

$$y = \sum_{i=1}^n w_i x_i + w_0 x_0 . \quad (5)$$

Тут x_i і w_i – відповідно сигнали на входах нейрона і ваги входів, функція u називається індукованим локальним полем, а $f(u)$ – передавальною функцією. Додатковий вхід x_0 і відповідну йому вагу w_0 використовують для ініціалізації нейрона.

Здатності нейронної мережі до прогнозування безпосередньо впливають з її здатності до узагальнення і виділення прихованих залежностей між вхідними та вихідними даними. Після навчання мережа здатна передбачити майбутнє значення якоїсь послідовності на основі декількох попередніх значень і (або) якихось існуючих зараз чинників. Перевагами цих моделей є можливість моделювання нелінійних процесів, масштабованість, висока адаптивність, різноманітність сфер застосування; недоліками – відсутність прозорості і неможливість аналізу отриманих результатів, складність вибору архітектури і алгоритму навчання, ресурсомісткість процесу навчання.

У цей час перспективним напрямком наукових досліджень вважають прогнозування з використання марковських моделей. Ідея застосування марковських моделей до прогнозування часових рядів ґрунтується на припущенні, що в основі коливань часового ряду лежить деякий марковський випадковий процес. Марковським називається випадковий процес, який має таку властивість: уся передісторія процесу повністю зосереджена у теперішньому стані, який тільки один впливає на майбутнє, і не важливо, як процес розвивався у минулому [7].

У випадку марковських моделей розглядається система, яка являє собою результат функціонування двох випадкових процесів. Перший – прихований процес, який спостерігач не може бачити, але цей процес безпосередньо впливає на другий процес – послідовність подій, яку бачить спостерігач.

На рис. 3 зображено схему марковської моделі. На схемі ми бачимо, що система переходить з одного прихованого стану в інший: $x(t-1)$, $x(t)$, $x(t+1)$... у той час, коли насправді ми спостерігаємо події $y(t-1)$, $y(t)$, $y(t+1)$.

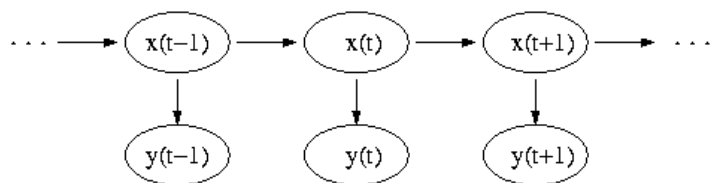


Рис. 3. Прихована марківська модель

Приховану марковську модель (ПММ) подають у вигляді $\lambda = \{S, \Omega, \pi, A, B\}$, де $S = \{s_1, \dots, s_N\}$ – множина станів системи; $\Omega = \{w_1, \dots, w_K\}$ – можливі події; $\pi = \{\pi_1, \dots, \pi_N\}$ – початкові ймовірності; $A = \{a_{ij}\}$ – матриця переходів; $B = \{b_{iw}\}$ – ймовірність спостереження події w_i після переходу до стану s_i [7].

Теорія марковських моделей не нова. Її основи опублікував Баум і його колеги в кінці 60-х - початку 70-х років минулого століття. Тоді ж на початку 70-х Бейкер і Джелінек з колегами з IBM застосували ПММ у розпізнаванні мови. Проте широко розповсюдженими ПММ стали нещодавно. За цей час вони показали досить якісні результати у таких областях прикладних досліджень, як

розпізнавання мови, біоінформатика, оброблення аудіо- та відеосигналів. В останній час зростає інтерес до використання цих моделей для вирішення задачі прогнозування часових рядів.

Переважає більшість досліджень в області прогнозування за допомогою марковських моделей присвячена проблемі прогнозування фінансових часових рядів. Це насамперед пов'язано з тим, що марковські моделі базуються на ймовірності подій, що дозволяє враховувати невизначеність і нелінійність, притаманну фондовим ринкам.

Порівняно з іншими моделями марковські моделі мають ряд переваг: їх легко будувати на експериментальних даних; вони не потребують розуміння внутрішніх механізмів динаміки змін системи; їх застосування не потребує великих витрат машинного часу; результати їх роботи можуть бути зрозумілі та пояснені з точки зору людини на відміну від інших методів; ці моделі мають сильне математичне обґрунтування. Результати експериментів показують, що моделі Маркова мають високий потенціал в області аналізу та прогнозування часових рядів.

В останній час дослідниками були запропоновані схеми застосування методу опорних векторів до розв'язання задачі прогнозування часового ряду. Основною ідеєю цього методу є переведення вихідних векторів в область більш високої розмірності та пошук поділяючої гіперплощини з максимальним зазором у цій області (рис.4).

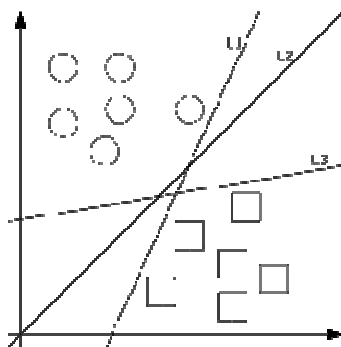


Рис.4. Побудова поділяючої гіперплощини

Слід зауважити, що таким чином можна побудувати прогноз тільки щодо майбутньої тенденції значень ряду, тобто зростання чи спадання показника. У рамках досліджень [8] вченими було виконано порівняння методу SVM з іншими методами прогнозування, а саме, лінійним дискримінантним аналізом (LDA), квадратичним дискримінантним аналізом (QDA), нейронними мережами зворотного розповсюдження Елмана (EBNN) та результатами, отриманими за допомогою моделі випадкового блукання (RW). Тестування проводилося на ряді, який являє собою тижневі коливання індексу NIKKEI 225. Найкращі результати були отримані з використанням підходу SVM і комбінованого підходу (Combining model), який являв собою середньозважене значень результатів, отриманих за всіма моделями.

Перевагами цього методу є математична обґрунтованість та зрозумілість методу. До недоліків можна віднести неможливість прогнозування значень рівнів ряду, тільки напрямку майбутнього тренду.

Зараз існує немало досліджень в області прогнозування часових рядів з використанням класифікаційно-регресійних дерев. Дерево прийняття рішень – це структура даних, у процесі обходу якої в кожному вузлі залежно від умови, яку перевіряють, приймають певне рішення про переміщення по тій чи іншій гілці дерева від кореня до кінцевої вершини. Алгоритми побудови класифікаційно-регресійних дерев відрізняються критерієм вибору атрибута, який потрібно покласти у вузлі на поточному кроці. Найбільш поширеними в області прогнозування часових рядів у цей час є алгоритми C4.5 та CART, тому що вони можуть працювати як з якісними, так і з кількісними даними. Згідно з роботою [9] прогнозне значення часового ряду залежить від попередніх рівнів ряду, а також деяких незалежних змінних. У прикладі, показаному на рис. 5, при визначенні значення цільової змінної враховуються як безперервні змінні, наприклад, $X(t)$, так і категоріальні Y , для яких виконується перевірка присутності значення в одному із заздалегідь визначених підмножин.

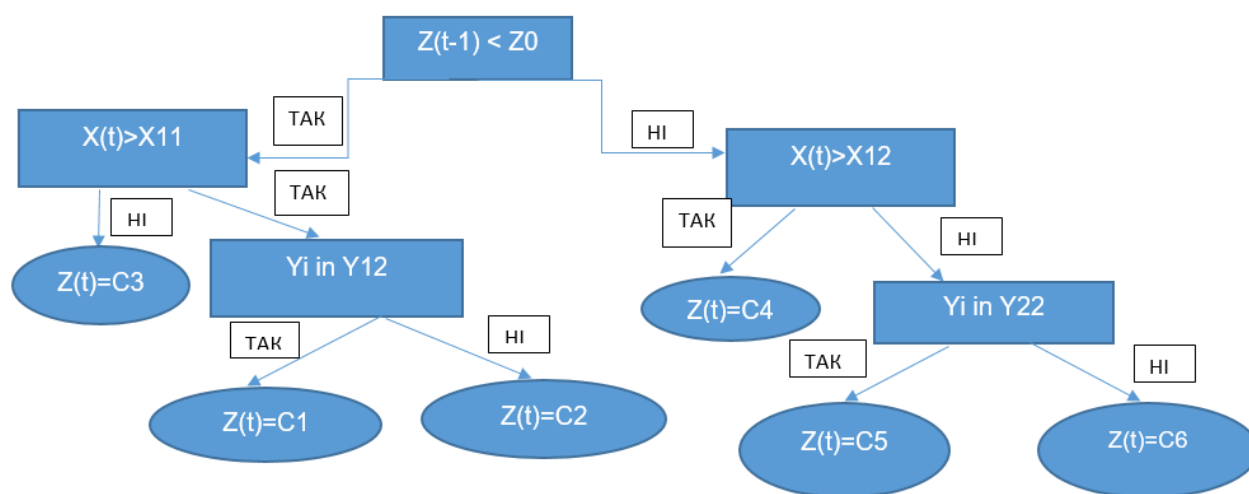


Рис. 5. Прогнозування часового ряду з використанням дерев рішень

Перевагами цих методів є те, що вони є простими у розумінні та інтерпретації, здатні працювати як з категоріальними, так і з неперервними змінними. Недоліками є проблема отримання найкращого дерева рішень для конкретного ряду. Ця задача є NP-повною з точки зору деяких аспектів оптимальності навіть для простих задач. Другою проблемою є проблема перенавчання дерева. Третя проблема полягає в тому, що ці алгоритми занадто чутливі до вихідних даних. Навіть невелика зміна в початковій послідовності може істотно змінити структуру дерева.

У роботі [10] наведені методи прогнозування часових рядів із використанням апарату нечіткої логіки. Прогнозування у випадку нечіткої логіки зводиться до того, щоб на основі вхідного часового ряду побудувати найбільш відповідну базу правил нечітких продукцій, тобто ту базу, з використанням якої, одержують найменшу похибку, та за допомогою цієї бази правил побудувати прогноз майбутніх значень ряду.

Базою правил нечітких продукцій називають кінцеву множину правил нечітких продукцій, узгоджених щодо застосовуваних у них лінгвістичних змінних. Найбільш часто базу правил подають у вигляді структурованого тексту [4]:

ПРАВИЛО_1: ЯКЩО «Умова_1», ТО «Висновок_1» (F_1).

ПРАВИЛО_2: ЯКЩО «Умова_2», ТО «Висновок_2» (F_2).

...

ПРАВИЛО_н: ЯКЩО «Умова_н», ТО «Висновок_н» (F_n).

Тут через $F_i, i=1, n$ позначено коефіцієнти визначеності, або вагові коефіцієнти відповідних правил. Ці коефіцієнти можуть набувати значень з інтервалу $[0, 1]$. У випадку, якщо вагові коефіцієнти відсутні, зручно вважати, що їх значення дорівнюють одиниці.

Математичною основою нечіткого моделювання часових рядів є теорема FAT (*Fuzzy Approximation Theorem*), згідно з якою будь-яка математична система може бути апроксимована системою, основаною на нечіткій логіці. Іншими словами, за допомогою природно-мовних висловлювань «ЯКЩО...,ТО» з подальшою їх формалізацією засобами теорії нечітких множин.

Головним недоліком цього методу є проблема знаходження оптимального розбиття універсуму та функцій належності значень ряду до певної нечіткої множини, що може вимагати участі аналітика і робить процес прогнозування за допомогою нечіткої логіки напіваавтоматичною процедурою.

Серед статистичних програмних пакетів для прогнозування і аналізу часових рядів можна виділити такі:

1. R – мова для статистичної обробки даних і роботи з графікою, а також вільне програмне середовище обчислень з відкритим вихідним кодом у рамках проекту GNU. Щодо аналізу часових рядів, R має бібліотеку *forecast*, яка містить реалізацію моделі прогнозування Хольта – Вінтерса та моделі ARIMA [11]. Перевагами пакета R є гнучкість, відкритий код і наявність потужних вбудованих аналітичних засобів. До недоліків відносять складність навчання для людей, які не є фахівцями в області математики та програмування.

2. MATLAB – пакет програм прикладних програм для розв'язання задач технічних розрахунків і назва мови програмування, що використовують у цьому пакеті. Для роботи з часовими рядами у середовищі Matlab вбудований пакет System Identification Toolbox. Він містить реалізацію лінійних і нелінійних моделей аналізу часових рядів, зокрема моделі AR, ARMAX, ARIMA, лінійну та нелінійну моделі ARX та інші [12]. Перевагами пакета є зручний графічний інтерфейс і простота в роботі, недоліками – дорогі ліцензії та досить заплутана інтеграція з JAVA і C ++ додатками.

3. STATISTICA – це інтегрована система аналізу і управління даними, а також інструмент розроблення додатків для бізнесу, економіки, фінансів, промисловості, медицини, страхування та інших областей. Система містить модуль для оброблення та аналізу часових рядів, який дозволяє проводити сезонну декомпозицію рядів динаміки, згладжування, кореляційний та спектральний аналіз, обчислення глобальних статистик часового ряду, а також прогнозування за допомогою моделі ARIMA та нейронних мереж [13]. До недоліків системи можна віднести те, що систему складно опановувати користувачам, які не є фахівцями в області математичної статистики.

4. Stata – це пакет статистичних прикладних програм, який містить модуль для оброблення та аналізу часових рядів, зокрема моделі прогнозування часових рядів ARIMA, ARCH / GARCH, реалізацію процедур Кохрейна-Оркутта, Прейс-Вінстена, обчислення оцінки Ньюї-Веста, побудову корелограми,

періодограми та інші [14]. Недоліками пакета є висока вартість, а також вузька спеціалізація в області економетрики, біометрії та аналізі стратифікованих обстежень.

5. Пакет SAS (Statistical Analysis System) – це професійний статистичний пакет, розроблений компанією SAS Institute Inc. Програма SAS містить реалізацію моделі Бокса-Дженкінса (ARIMA), регресійних моделей та векторних авторегресійних моделей прогнозування часових рядів, експоненціальне згладжування (ESM), спектральний (SPECTRA) і фазовий (STATESPACE) аналіз часових рядів [15]. Перевагами SAS є швидке оброблення дуже великих обсягів даних; можливість перетворення математичних формул у програмний код; створення призначених для користувача модулів; гнучкий інтерфейс обміну даними, а також наявність інструментарію для роботи з кластерами (розподіленими системами). Недоліками є дорога вартість ліцензій та складність освоєння статистичного пакета для новачків.

6. SPSS (Statistical Package for the Social Sciences) – комп'ютерна програма для статистичної обробки даних, один із лідерів ринку в області комерційних статистичних продуктів, призначених для проведення прикладних досліджень у соціальних науках. Програма SPSS містить реалізацію моделі прогнозування часових рядів ARIMA, пошук аномальних значень ряду, моделі експоненціального згладжування часових рядів [16]. Перевагами пакета є розвинений апарат статистичного аналізу; універсальність; широкий набір статистичних і графічних процедур аналізу даних; висока швидкість обчислень; простий і зручний інтерфейс. Недоліками SPSS є те, що її використання накладає високі вимоги до системи комп'ютера, та висока ціна в порівняно зі статистичними пакетами аналогічного рівня.

7. Sales-Forecast – це програмна система, яка дозволяє в автоматичному режимі будувати прогнози різного ступеня деталізації для великого числа товарів і товарних груп багатонаменклатурної продукції, враховуючи при цьому календар подій і індивідуальні особливості часових рядів. У Sales-Forecast втілені відомі методи прогнозування: методи на основі моделей ARIMA та експоненціального згладжування, методи сезонної декомпозиції, спектрального аналізу, побудови профілів продажів, а також за допомогою нейронних мереж [17]. Перевагами цієї системи є те, що вона проста у використанні й не потребує суттєвих знань в області математичної статистики від користувача. Недоліками є висока вартість і вузька спеціалізація, а саме зосередженість на часових рядах, які являють собою рівень об'ємів продаж.

8. Forecast Pro – програмне забезпечення для побудови прогнозів, орієнтоване на бізнес-застосування. Пакет Forecast Pro забезпечує автоматичний підбір прогнозової моделі з такого набору: прості моделі експоненціального згладжування; моделі експоненціального згладжування Хольта; сезонні моделі експоненціального згладжування Вінтерса; авторегресійні моделі Бокса-Дженкінса, включаючи адитивну і мультиплікативну модель; моделі переривчастого попиту Кростона; модель Census X11; моделі дискретного попиту; моделі динамічних регресій; найпростіші моделі змінного середнього; класичні криві тренду (лінійне, квадратичне, експоненційне зростання) [18]. Перевагою системи є її зосередженість на розв'язку задачі прогнозування часових рядів, простота у використанні, широкий набір реалізованих методів прогнозування. Недоліком системи є її висока вартість.

9. KXEN – це інструмент для моделювання, в основі якого лежить підхід Data Mining. Пакет KXEN містить компонент аналізу часових рядів KXEN Time Series - KTS, який дозволяє виділяти значущі шаблони і тренди в часових рядах, а також періодичну, сезонну складову та залишки. Програма KXEN дозволяє будувати короткі та середньострокові прогнози [19].

10. Weka (Waikato Environment for Knowledge Analysis) – безкоштовне програмне забезпечення для аналізу даних і машинного навчання. Версія Weka 3.7.3 містить розширений модуль для аналізу та прогнозування часових рядів, який містить створення лагів, виділення періодичної складової, побудову моделі ряду, її оцінку та оцінку інтервалів довіри прогнозів [20].

11. Zaitun Time Series являє собою практичний інструмент, який можна використовувати для статичного аналізу, моделювання та прогнозування даних часових рядів. Пакет містить модулі для обчислення статистик ряду, побудови нейронних мереж, а також графічні інструменти, які полегшують аналіз часових рядів. Програма дозволяє аналізувати наявність існуючих у ряді тенденцій, проводити його декомпозицію, згладжування за допомогою ковзкого середнього та експоненціальне згладжування, будувати корелограму [21]. Окрім цього, Zaitun Time Series має можливість роботи з даними ринку.

12. NCSS – це комп'ютерна програма для аналізу статистичних даних. NCSS містить 230 задокументованих статистичних процедур для аналізу даних і прогнозування. Програма NCSS містить модель прогнозування Бокса-Дженкінса ARIMA, модель ARMA, обчислення автокореляції, декомпозицію часового ряду на складові частини, експоненціальне згладжування, гармонічний аналіз, лагову трансформацію, модифіковані рівняння Юла-Уолкера, автокореляцію залишків, аналіз сезонної та трендової складової, спектральний аналіз [22].

13. GMDH Shell – це програмний інструмент для інтелектуального аналізу даних і прогнозування на основі МГУА (методу групового урахування аргументів). Програмний пакет дає можливість переглянути формулу побудованої моделі і отримані прогнози. Він дозволяє працювати з різними типами електронних таблиць і баз даних. Результати роботи подають у вигляді зручних таблиць і графіків [23].

14. UNISTAT – програмний пакет для статистичного аналізу даних. У програмі реалізовано модель прогнозування часових рядів ARIMA, експоненціальне згладжування, адаптивні моделі прогнозування часових рядів Брауна, Хольта та Вінтерса, тест тренду Ньюманна, регресію Кокса, аналіз Каплана-Мейера, аналіз Вейбулла, побудову діаграм Паретто, пряму та зворотну трансформації Фур'є [24]. Недоліком цього пакета, а також GMDH Shell та NCSS є висока вартість ліцензій.

Висновки

У роботі був проведений огляд методів, моделей та програмних засобів прогнозування часових рядів. У зв'язку з тим, що кількість методів і моделей прогнозування сьогодні стала дуже великою, виникає необхідність класифікації існуючих методів і моделей прогнозування. Методи прогнозування поділяють на інтуїтивні, тобто ті, які мають справу із судженнями експертів, та формалізовані – мають справу з математичними моделями. Серед моделей прогнозування виділяють моделі предметної області та моделі часових рядів. Моделі часових рядів, у свою чергу, поділяють на статистичні та структурні. До статистичних моделей відносять регресійні моделі, моделі ковзкого середнього, моделі ARIMA, метод групового урахування аргументів та інші. До структурних моделей відносять

моделі, які ґрунтуються здебільшого на принципі машинного навчання та намагаються знайти деякі закономірності розвитку процесу всередині самого часового ряду, наприклад, моделі на основі нечіткої логіки, дерев рішень, нейронних мереж, ланцюгів Маркова, генетичних алгоритмів, опорних векторів тощо. Сьогодні перспективним напрямком досліджень вважається прогнозування часових рядів із використанням структурних моделей.

Аналіз програмних продуктів для прогнозування часових рядів показав, що їх можна поділити на такі групи: статистичні пакети, які мають велику кількість реалізованих функцій і методів з аналізу даних, такі пакети містять модуль з прогнозування часових рядів (наприклад, R, MATLAB, STATISTICA, NCSS, Stata, SAS) і пакети, спеціалізовані на прогнозуванні часових рядів, такі, як Zaitun Time Series, Forecast Pro, Sales Forecast. Певні пакети (Sales Forecast, Forecast Pro, UNISTAT, GMDH Shell, SPSS) є простими у використанні, навіть для людей, які не є фахівцями в області аналізу даних, але мають дорогі ліцензії. Інші, наприклад R, є безкоштовними, але вимагають від користувача знань в області програмування та статистики. Існують також безкоштовні пакети, які є простими в освоєнні та зрозумілими у використанні, наприклад Zaitun Time Series, але містять обмежену кількість методів прогнозування часових рядів. Крім цього, незважаючи на зростаючий інтерес до структурних моделей прогнозування, кількість програмних пакетів, які містять їхню реалізацію, є обмеженою. Тому виникає необхідність створення програмного продукту, який би містив реалізацію багатьох моделей прогнозування, як статистичних, так і структурних; надавав би можливість порівняння результатів прогнозування, отриманих за допомогою цих методів; був би простим в освоєнні не тільки для спеціалістів в області математичної статистики та програмування, але й для звичайних користувачів.

Особливу увагу в процесі подальших досліджень планується приділити марковським моделям. Ці моделі зарекомендували себе як дуже потужний засіб прогнозування фінансових часових рядів, які характеризуються нелінійністю та невизначеністю. Однак залишаються відкритими такі питання, як проблема вибору оптимального розміру тренувальної послідовності, вибір функцій розподілу для обчислення значень емісійної матриці тощо. Більш детальний опис застосування марковських моделей для вирішення задачі прогнозування часового ряду та нерозв'язаних проблем в цій області проведено у роботі [25].

Список літератури

1. Чучуева, И.А. Модель прогнозирования временных рядов по выборке максимального правдоподобия: дис. канд. техн. наук / Чучуева И.А. - // Московский гос. техн. ун-т им. Н.Э. Баумана. - Москва, 2012.
2. Тихонов, Е.Е. Методи прогнозування в умовах ринку : навч. посібник / Е. Е. Тихонов.- // Невинномиськ, 2006. – 221с
3. Jingfei Yang M. Sc. Power System Short-term Load Forecasting: Thesis for Ph.d degree. // Germany, Darmstadt, Elektrotechnik und Informationstechnik der Technischen Universitat, 2006. 139 p.
4. Лукашин Ю.П. Адаптивні методи короткострокового прогнозування часових рядів : навч. посібник // Ю.П. Лукашин. - М.: Фінанси і статистика, 2003. – 416 с.
5. Autoregressive conditional heteroskedasticity [Електронний ресурс]. — Режим доступу:
https://en.wikipedia.org/wiki/Autoregressive_conditional_heteroskedasticity

6. Нейронні мережі [Електронний ресурс]. — Режим доступу: <http://www.aiportal.ru/articles/neural-networks/neural-networks.html>
7. Erhard Rank, Franz Pernkopf. Hidden Markov Models. // Lecture Notes Speech Communication 2, SS 2004, Signal Processing and Speech Communication Laboratory, Graz University of Technology.
8. Wei Huangab, Yoshiteru Nakamoria, Shou-Yang Wangb. Forecasting stock market movement direction with support vector machine // Computers & Operations Research 32. – 2005. –P. 2513 – 2522.
9. Hannes Y.Y., Webb P. Classification and regression trees: A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity // International Food Policy Research Institute. – 1999. –P. 59.
10. Shyi Ming Chen, Nien Yi Chung. Forecasting Enrollments of Students by Using Fuzzy Time Series and Genetic Algorithms // Information and management sciences: Volume 17, Number 3, -P. 1-17, 2006.
11. CRAN - Contributed Packages [Електронний ресурс].— Режим доступу: <https://cran.r-project.org/web/packages/>
12. System identification toolbox [Електронний ресурс].— Режим доступу: <https://www.mathworks.com/products/sysid.html>
13. Методи прогнозування [Електронний ресурс].— Режим доступу: <http://statsoft.ru/solutions/tasks/forecast/>
14. Stata 7 [Електронний ресурс].—
Режим доступу: <http://old.exponenta.ru/soft/others/stata/stata.asp>
15. Прогнозування та аналіз часових рядів. Лекція 1. SAS. [Електронний ресурс].—
Режим доступу: https://www.sas.com/content/dam/SAS/ru_ru/doc/academic/VMK_MGU/2015/lec7/EM2015_7.pdf
16. Time Series Modeler. [Електронний ресурс].— Режим доступу: https://www.ibm.com/support/knowledgecenter/en/SSLVMB_22.0.0/com.ibm.spss.statistics.help/spss/trends/idh_idd_tab_vars.htm
17. Эффективные инструменты прогнозирования продаж: система Sales-Forecast. [Електронний ресурс].— Режим доступу: <http://softline.ru/news/22529>
18. Forecast Pro – программное обеспечение для бизнес-прогнозирования. [Електронний ресурс]. Переглянуто 9 листопада 2017. <http://prospo.ru/winmat/2285-forecast-pro->
19. Kxen.TimeSeries. [Електронний ресурс]. — Режим доступу: <https://help.sap.com/viewer/e298bfb935ae49999de48dc9c269b90b/3.2/en-US/7d2350147372101483e3e699b0e91070.html>
20. Weka. [Електронний ресурс].—
Режим доступу: <https://ru.wikipedia.org/wiki/Weka>
21. Zaitun Time Series Online Documentation. [Електронний ресурс].—
Режим доступу: <http://www.zaitunsoftware.com/content/zaitun-time-series-online-documentation>
22. NCSS (statistical software). [Електронний ресурс].— Режим доступу [https://en.wikipedia.org/wiki/NCSS_\(statistical_software\)](https://en.wikipedia.org/wiki/NCSS_(statistical_software))
23. Time Series Analysis Solution for Business. [Електронний ресурс].—
Режим доступу <https://www.gmdhshell.com/time-series-analysis-software>
24. Forecasting and Smoothing. [Електронний ресурс].— Режим доступу: <https://www.unistat.com/guide/time-series-analysis-forecasting-and-smoothing/>

25. Долгіх, А.О. Огляд сучасних розробок прогнозування часових рядів з використанням прихованих марківських моделей. / А.О. Долгіх, О.Г. Байбуз. // Актуальні проблеми автоматизації та інформаційних технологій : зб. наук. пр. – Дніпро : Ліра, 2017. – Т. 21.

Надійшла до редколегії 13.02.2018

Анализ методов, моделей и программных средств прогнозирования временных рядов

Проведен анализ существующих в настоящее время методов и моделей прогнозирования временных рядов и дана их классификация. Представлено описание как классических статистических методов и моделей прогнозирования, такие, как модели экспоненциального сглаживания, ARIMA модели, регрессионные модели, так и более современных, основанных на принципе машинного обучения, например, модель на основе опорных векторов (SVM), нейронные сети, марковские модели и классификационно-регрессионные деревья. Описаны программные пакеты, которые содержат модули для анализа и прогнозирования временных рядов. Представлены направления будущих исследований и возможных программных разработок в области прогнозирования финансовых временных рядов.

Ключевые слова: классификация методов и моделей прогнозирования временных рядов, адаптивные модели, ARIMA модели, регрессионные модели, SVM, марковские модели, нейронные сети, классификационно-регрессионные деревья.

The Analysis of Time Series Forecasting Methods, Models and Software Packages

The analysis of existing for today's day time series forecasting methods and models and their classification is carried out. The description of both classical statistical forecasting methods and models, such as exponential smoothing models, ARIMA models, regression models, and more modern ones based on the principle of machine learning, for example, the model based on support vectors (SVM), neural networks, markov models and classification-regression trees is presented. Software packages that contain modules of time series analysis and forecasting are described. The directions of future research and possible software developments in the field of financial time series forecasting are presented.

Key words: classification of methods and models of time series forecasting, adaptive models, ARIMA models, regression models, SVM, markov models, neural networks, decision trees.

Відомості про авторів:

Долгіх Анастасія Олегівна – аспірантка кафедри математичного забезпечення електронно-обчислювальних машин; Дніпровський національний університет ім. Олеся Гончара; Україна.

Байбуз Олег Григорович – доктор технічних наук, професор, завідувач кафедрою математичного забезпечення електронно-обчислювальних машин; Дніпровський національний університет ім. Олеся Гончара; Україна.