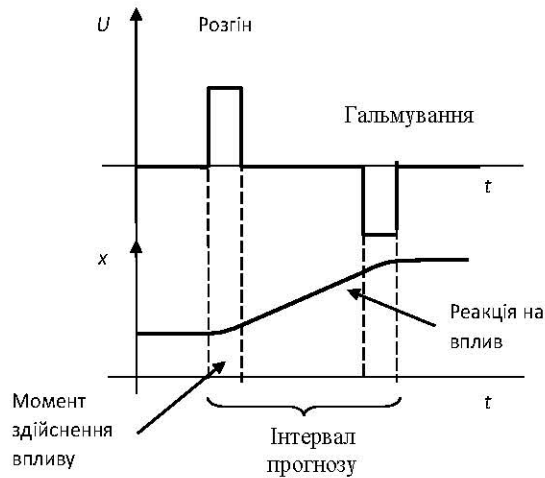


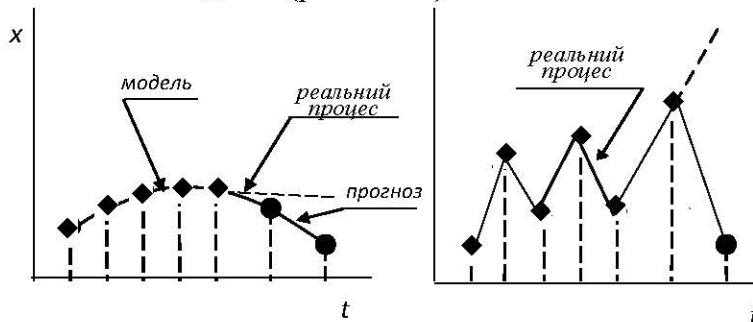
12.2. Використання моделі для оптимального прогнозування

12.2.1. Класифікація задач і методів прогнозування

Прогнозування стану об'єкта керування є одним з найважливіших застосувань моделі в системах керувань. Адже *інерційність процесів* керування приводить до того, що реакція об'єкта на керуючий вплив відбувається через певний час після здійснення впливу (рис. 12.9). Таким чином для здійснення правильного впливу необхідно передбачити стан об'єкта та всіх незалежно зовнішніх впливів.



Прогнозування може здійснюватись як на основі детермінованої, так і на основі стохастичної моделей (рис. 12.10).



Детерміноване прогнозування використовується при незначних випадкових впливах і невеликій кількості даних про попередній хід процесу.

При *стохастичному прогнозуванні* хід прогнозованого процесу у часі за попередній період необхідно подати функцією $X(t)$, що отримується за допомогою *методів стохастичної ідентифікації*. Для задачі прогнозування найчастіше використовується регресійна модель. Нагадаємо, що загальний вигляд рівняння регресії (6.26)

$$y(t) = \bar{y}(t) + \varepsilon,$$

де $y(t)$ – реальне значення випадкового процесу; $\bar{y}(t)$ – значення випадкового процесу, отримане за допомогою функції регресії; ε – випадкова величина, що характеризує вплив неврахованих факторів, найчастіше – білий шум.

З математичної точки зору детерміноване прогнозування є *екстраполяцією* математичної моделі керованого процесу.

Задача екстраполяції полягає у визначенні значення функції $y(x)$ в точці, що не належить відрізьку $[x_0, x_n]$, на якому визначені вузли екстраполяції. Для екстраполяції використовуються ті ж вирази, що і при інтерполяції, але з більшою похибкою. Так, для гладких функцій екстраполяція доцільна при значеннях x , що виходять за зазначені межі не більш ніж на $h/2$, де h – крок розташування вузлів.

Алгоритм екстраполяції залежить від вибору функцій екстраполяції. Але у будь-якому випадку для екстраполяції n -го порядку необхідно зберігати n даних у буфері зі структурою черги.

12.2.2. Основи прогнозування даних

Розглянемо основи прогнозування даних на прикладі розв'язання задачі короткострокового прогнозування значень часового ряду методом експоненційного згладжування (the Method of Moving Average за Р. Брауном). Також розглянемо питання структурного моделювання систем прогнозування на основі таких алгоритмів агрегування моделей (створення ансамблів моделей), як бегінг (bagging), бустинг (boosting) і стекінг (stacking).

12.2.2.1 Часові ряди і стохастичні процеси

Розглянемо лише дискретні часові ряди, в яких спостереження проводяться через фіксований інтервал часу, що приймається за одиницю відліку. Перехід від моменту одного спостереження до моменту наступного спостереження називатимемо кроком.

Відомо, що якщо значення членів часового ряду точно визначені якою-небудь математичною функцією, то часовий ряд називається детермінованим. Якщо ці значення можуть бути описані лише за допомогою розподілу ймовірностей, часовий ряд називається випадковим.

Явище, яке розвивається у часі випадковим чином, можна розглядати й називати стохастичним процесом у випадку, якщо механізм випадкості можна

описати в поняттях теорії ймовірності. В іншому випадку випадковий процес необхідно розглядати як хаотичний.

В даному розділі розглядаються лише стохастичні процеси. При цьому відрізок часового ряду, що аналізується, розглядається як одна з окремих реалізацій (вибірка) стохастичного процесу, що вивчається, з прихованим ймовірнісним механізмом.

В свою чергу, серед стохастичних процесів виділяють клас процесів, які називають стаціонарними. Для їх математичної формалізації позначимо член часового ряду, що спостерігається в момент t , через x_t . Стохастичний процес називається стаціонарним, якщо його властивості не змінюються в часі. Зокрема, він має постійне математичне очікування $\bar{x} = M(x_t)$ (середнє значення, відносно якого він варіює), постійну дисперсію $D(x) = M[(x_t - \bar{x})^2] = \sigma_x^2$, що визначає розмах його коливань відносно середнього значення, а також постійну автоковаріацію й коефіцієнти автокореляції.

Коваріація між значеннями x_t і x_{t+k} , які відділені інтервалом в k одиниць часу, називається автоковаріацією з лагом (затримкою) k і визначається як

$$R_{xx}(k) = \text{cov}(x_t, x_{t+k}) = M[(x_t - \bar{x})(x_{t+k} - \bar{x})].$$

Для стаціонарних процесів автоковаріація залежить лише від лагу k й $R_{xx}(0) = \sigma_x^2$. Автокореляція з лагом k є лише нормованою автоковаріацією і дорівнює:

$$p_k = \frac{M[(x_t - \bar{x})(x_{t+k} - \bar{x})]}{\sqrt{M[(x_t - \bar{x})^2]} \sqrt{M[(x_{t+k} - \bar{x})^2]}} = \frac{M[(x_t - \bar{x})(x_{t+k} - x)]}{\sigma_x^2},$$

оскільки для стаціонарного процесу $\sigma_x^2 = \text{const}$. Таким чином, k -й коефіцієнт автокореляції

$$p_k = \frac{R_{xx}(k)}{R_{xx}(0)}.$$

Коефіцієнт автокореляції має таку властивість, що $-1 \leq p_k \leq 1$.

Уявимо, що часовий ряд який генерується деякою системою, можна подати у вигляді двох компонент:

$$x_t = \xi_t + \varepsilon_t,$$

де величина ξ_t – випадковий неавтокорельований процес з нульовим математичним очікуванням й кінцевою (не обов'язково постійною) дисперсією, а величина ε_t може бути або детермінованою функцією, або випадковим процесом, або якою-небудь їх комбінацією. Величини ε_t і ξ_t розрізняються харак-

тером впливу на значення наступних членів ряду. Змінна ε_t впливає лише на значення синхронного їй члену ряду, в той час як величина ξ_t у відомому степені визначає значення декількох чи всіх наступних членів ряду. Через величину ξ_t здійснюється взаємодія членів ряду; таким чином, в ній міститься інформація про систему, необхідна для отримання прогнозів. Назвемо величину ξ_t рівнем ряду в момент t , а закон еволюції рівня в часі – *трендом*. Таким чином, тренд може бути виражений як детермінованою, так і випадковою функціями, або їх комбінацією. Стохастичні тренди мають, наприклад, ряди з випадковим рівнем або випадковим стрибкоподібним характером зміни (випадкові процеси хвильової структури).

Наведемо приклад детермінованого тренду 2-го порядку:

$$\xi_t = a_1 + a_2 t + a_3 t^2,$$

де a_1, a_2, a_3 — постійні коефіцієнти; t — час.

Приклад випадкового тренду:

$$\xi_t = \xi_{t-1} + u_t = \xi_0 + \sum_{i=1}^t u_i$$

де ξ_0 — початкове значення; u_t — випадкова змінна.

Приклад тренду змішаного типу:

$$\xi_t = a_1 + a_2 t + u_t + q u_{t-1} + b \sin \omega t,$$

де a_1, a_2, q, b, ω - постійні коефіцієнти; u_t — випадкова змінна.

Відомо велику кількість визначень рівня і тренду ряду (див. [91, с.16]). Існуючі поняття тренду суперечливі і носять умовний характер. Кожне з цих означень скоріше вказує на окремий спосіб оцінювання тренду, а не на його сутність. Дуже часто під трендом розуміють детерміновану складову процесу, що значно збіднює зміст терміна й перешкоджає його використанню для аналізу часових рядів в загальному випадку.

Компоненти часового ряду ξ_t і ε_t неспостережні. Вони являють собою теоретичні величини. Саме їх виділення і складає предмет аналізу часового ряду в задачі прогнозування. Оцінюють майбутніх членів ряду зазвичай роблять за прогнозною моделлю. Прогнозна модель — це модель, яка апроксимує тренд. Прогнози — це оцінки майбутніх рівнів ряду, а послідовність прогнозів для різних періодів випередження $\tau = 1, 2, \dots, k$ складає оцінку тренду.

При побудові прогнозної моделі висувається гіпотеза про динаміку величини ξ_t , тобто про характер тренду. Однак у зв'язку з тим, що впевненість у гіпотезі завжди відносна, моделі, що розглядаються, наділяються адаптивними властивостями, здатністю до корекції вихідної гіпотези або навіть до заміни її

іншою, яка більш адекватно (з точки зору точності прогнозів) відображає поведінку реального ряду.

Найпростіша адаптивна модель базується на обчисленні так званого експоненційного середнього (за Р. Брауном – ковзного середнього (moving average)).

12.2.2.2 Експоненційне згладжування

Припустимо, що досліджується часовий ряд x_t . Виявлення й аналіз тенденції динамічного ряду часто здійснюється за допомогою його вирівнювання або згладжування. Експоненційне згладжування – один з найпростіших і розповсюджених прийомів вирівнювання ряду. В його основі лежить розрахунок експоненційних середніх.

Експоненційне згладжування ряду здійснюється по рекурентній формулі, яка по суті описує марковський ітераційний процес апроксимації вихідного ряду x_t :

$$S_t = \alpha x_t + \beta S_{t-1}, \quad (12.2)$$

де S_t — значення експоненційної середньої в момент t ;

α — параметр згладжування, $\alpha = const, 0 < \alpha < 1; \beta = 1 - \alpha$.

Вираз (12.2) можна переписати таким чином:

$$S_t = \alpha x_t + (1 - \alpha)S_{t-1} = S_{t-1} + \alpha(x_t - S_{t-1}).$$

Експоненційне середнє на момент t тут виражене як експоненційне середнє попереднього моменту плюс частка α -різниці поточного спостереження й експоненційного середнього минулого моменту (moving average).

Якщо послідовно використовувати це рекурентне співвідношення, то експоненційне середнє S_t можна виразити через значення часового ряду x :

$$\begin{aligned} S_t &= \alpha x_t + \beta S_{t-1} = \alpha x_t + \alpha \beta x_{t-1} + \beta^2 S_{t-2} \dots = \\ &= \alpha x_t + \alpha \beta x_{t-1} + \alpha \beta^2 x_{t-2} + \dots + \alpha \beta^i x_{t-i} + \dots + \beta^N S_0 = \alpha \sum_{i=0}^{N-1} \beta^i x_{t-i} + \beta^N S_0, \end{aligned} \quad (12.3)$$

де N — кількість членів ряду; S_0 — деяка величина, що характеризує початкові умови для першого використання формули (12.2) при $t=1$.

Оскільки $\beta < 1$, то при $N \rightarrow \infty, \beta^N \rightarrow 0$, сума коефіцієнтів $\alpha \sum_{i=0}^{N-1} \beta^i \rightarrow 1$.

Тоді

$$S^i = \alpha \sum_{i=0}^{\infty} \beta^i x_{t-1}.$$

Таким чином, величина S_t є зваженою сумою всіх членів ряду. Причому ваги падають експоненційно залежно від давності («віку») спостереження. Це і

пояснює, чому величина S_t названа експоненційним середнім. Якщо, наприклад, $\alpha = 0.3$, то поточне спостереження буде мати вагу 0.3, а ваги попередніх даних складуть відповідно 0,21; 0,147; 0,103 і т.д.

Розглянемо ряд, згенерований моделлю

$$x_t = a_1 + \varepsilon_t,$$

де $a_1 = \text{const}$; ε_t – випадкові неавтокорельовані відхилення, або шум з середнім значенням 0 і дисперсією σ^2 .

Застосуємо до нього процедуру експоненційного згладжування (12.2). Тоді

$$S_t = \alpha \sum_{i=0}^{\infty} \beta^i x_{t-1} = \alpha \sum_{i=0}^{\infty} \beta^i (a_1 + \varepsilon_{t-1}) = a_1 + \alpha \sum_{i=0}^{\infty} \beta^i \varepsilon_{t-1}.$$

Знайдемо математичне очікування

$$M(S_t) = M(x_t) = a_1$$

і дисперсію

$$D(S_t) = M[(S_t - a_1)^2] = M[(\alpha \sum_{i=0}^{\infty} \beta^i \varepsilon_{t-1})^2] = \alpha^2 \sum_{i=0}^{\infty} \beta^{2i} \sigma^2 = \frac{\alpha}{2 - \alpha} \sigma^2. \quad (12.4)$$

Оскільки $0 < \alpha < 1$, то $D(S_t) < D(x_t) = \sigma^2$.

Таким чином, експоненційне середнє S_t має те ж математичне очікування, що й ряд x , але меншу дисперсію. Як видно з (12.4), при високому значенні α дисперсія експоненційного середнього несуттєво відрізняється від дисперсії ряду x_t . Чим менше α , тим більшою мірою скорочується дисперсія експоненційного середнього. Відповідно, експоненційне згладжування можна розглядати як фільтр, на вхід якого у вигляді потоку послідовно надходять члени вихідного ряду, а на виході формуються поточні значення експоненційного середнього. І чим менше α , тим більшою мірою фільтруються, приглушуються коливання вихідного ряду.

Після появи робіт Р. Брауна експоненційне середнє часто використовується для короткострокового прогнозування. У цьому випадку припускається, що ряд генерується моделлю

$$x_t = a_{1,t} + \varepsilon_t,$$

де $a_{1,t}$ – середній рівень ряду, що варіюється в часі; ε_t – випадкові неавтокорельовані відхилення з нульовим математичним очікуванням і дисперсією σ^2 .

Прогнозна модель (предиктор) має вигляд

$$\hat{x}_t(t) = \hat{a}_{1,t},$$

де $\hat{x}_t(t)$ — прогноз, зроблений в момент t на τ одиниць часу (кроків) вперед;

$\hat{a}_{1,t}$ — оцінка $a_{1,t}$ (знак \wedge над величиною тут і далі означатиме оцінку).

Оцінкою єдиного параметра моделі слугує експоненційне середнє. $\hat{a}_{1,t} = S$ Таким чином, усі властивості експоненційного середнього поширюються на прогнозну модель. Зокрема, якщо S_{t-1} розглядати як прогноз на 1 крок вперед, то у виразі (12.2) величина $x_t - S_{t-1}$ являє собою похибку цього прогнозу, а новий прогноз S_t отримується в результаті корекції попереднього прогнозу з врахуванням його помилки. В цьому і полягає сутність адаптації як процесу уточнення наступної точки прогнозу на основі поточної інформації про його попереднє значення.

При короткостроковому прогнозуванні бажано якнайшвидше відобразити зміни $a_{1,t}$ і в той же час якнайкраще «очистити» ряд від випадкових коливань.

Таким чином, з одного боку, варто збільшувати вагу більш свіжих спостережень, чого можна досягнути підвищенням α (12.3), з іншого боку, для згладжування випадкових відхилень величину α варто зменшити. Як бачимо, ці дві вимоги суперечать одна одній. Пошук компромісного значення α складає задачу оптимізації моделі.

Для з'ясування процедури розрахунку експоненційного середнього і його властивостей розглянемо числовий приклад згладжування ряду курсу акцій фірми IBM (табл. 12.1).

Результати обчислень експоненційних середніх при $\alpha = 0.1$, $\alpha = 0.5$ і $\alpha = 0.9$ наведені в табл. 12.1.

Таблиця 12.1 – Експоненційні середні

№ точки (час)	Члени ряду				№ точки (час)	Члени ряду			
		$\alpha=0.1$	$\alpha=0.5$	$\alpha=0.9$			$\alpha=0.1$	$\alpha=0.5$	$\alpha=0.9$
1	510	506,4	508,0	509,6	16	512	505,7	513,3	513,1
2	497	505,5	502,5	498,3	17	510	506,1	511,7	510,3
3	504	505,3	503,2	503,4	18	506	506,1	508,8	506,4

4	510	505,8	506,6	509,3	19	515	507,0	511,9	514,1
5	509	506,1	507,8	509,0	20	522	508,5	517,0	521,2
6	503	505,8	505,4	503,6	21	523	509,9	520,0	522,8
7	500	505,2	502,7	500,4	22	527	511,6	523,5	526,6
8	500	504,7	501,4	500,0	23	523	512,8	523,2	523,4
9	500	504,2	500,7	500,0	24	528	514,3	525,6	527,5
10	495	503,3	497,8	495,5	25	529	515,8	527,3	528,9
11	494	502,4	495,9	494,2	26	538	518,0	532,7	537,1
12	499	502,0	497,5	498,5	27	539	520,1	525,8	538,8
13	502	502,0	499,7	501,2	28	541	522,2	538,4	540,8
14	509	502,7	504,4	508,3	29	543	524,3	540,7	542,8
15	525	505,0	514,7	523,3	30	541	525,9	540,9	541,2

Визначимо S_0 як $\frac{1}{5} \sum_{i=0}^5 x_i = \frac{1}{5}(510 + 497 + 504 + 510 + 509) = 506$.

Подальші обчислення при $\alpha = 0.1$ виглядають так:

$$S_1 = \alpha x_1 + (1 - \alpha)S_0 = 0.1 \times 510 + 0.9 \times 506 = 506.4;$$

$$S_2 = \alpha x_2 + (1 - \alpha)S_1 = 0.1 \times 497 + 0.9 \times 506.4 = 505.46;$$

$$S_3 = \alpha x_3 + (1 - \alpha)S_2 = 0.1 \times 504 + 0.9 \times 505.46 = 505.31$$

і т. д.

На рис. 12.11 зображений графік динаміки часового ряду й експоненційних середніх при $\alpha = 0.1$ і $\alpha = 0.5$. На графіку наочно проявляється вплив величини α на рухливість експоненційного середнього.

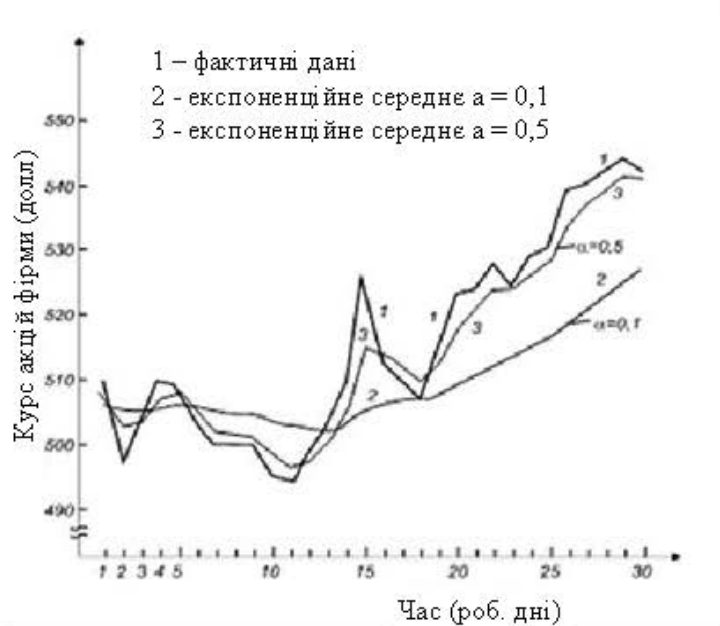


Рисунок 12.11 – Експоненційне згладжування часового ряду

Експоненційне згладжування являє собою найпростіший варіант самонавчальної моделі. Обчислення прості і виконуються ітеративно. Вони потребують навіть менше арифметичних операцій, ніж ковзне середнє, а масив минулої інформації зменшений до одного значення S_{t-1} . Таку модель будемо називати адаптивною експоненційного типу, а величину α – параметром адаптації. Дослідімо її властивості.

12.2.2.3 Початкові умови експоненційного згладжування

Експоненційне вирівнювання завжди потребує попереднє значення експоненційного середнього. Коли процес тільки починається, повинна бути певна величина S_0 , яка може бути використана як значення, що передувало S_1 . Якщо є минулі дані до моменту початку вирівнювання, то за початкове значення S_0 можна використовувати арифметичне середнє всіх наявних точок або якої-небудь їх частини. Коли для такого оцінювання S_0 немає даних, необхідне передбачення початкового рівня ряду.

Передбачення може бути зроблене, виходячи з апріорних знань про процес або на основі його аналогії з іншими процесами. Після k кроків вага, надана початковому значенню, дорівнює $(1 - \alpha)^k$. Якщо є впевненість в справедливості початкового значення S_0 , то можна коефіцієнт α взяти малим. Якщо такої впевненості немає, то параметру α варто надати більше значення, з таким розрахунком, щоб вплив початкового значення швидко зменшився. Однак велике значен-

ня α , як це впливає з (12.4), може стати причиною великої дисперсії коливань S_t . Якщо необхідно згасити ці коливання, то після достатнього віддалення від початкового моменту часу величину α можна зменшити.

Розглянемо роль параметра α в початковий період згладжування у випадку, коли немає впевненості в справедливості вибору початкової величини S_0 .

Таблиця 12.2 – Зміна ваг в початковий період часу при експоненційному згладжуванні з $\alpha=0,1$

Ітерація	Вага початкової величини	Вага першого члена ряду	Вага другого члена ряду	Вага третього члена ряду	Вага четвертого члена ряду
1	0,900	0,100			
2	0,810	0,090	0,100		
3	0,729	0,081	0,090	0,100	
4	0,656	0,073	0,081	0,090	0,100

Початкова величина S_0 , при збільшенні кількості ітерацій, протягом довгого часу має надмірну вагу. Навіть після 20 ітерацій вага S_0 дорівнює 0,122, що означає, що йому дається все ще більша вага, ніж будь-якому іншому члену ряду. Таким чином, у цьому випадку отримання прогнозів за експоненційним середнім, що побудоване на малому відрізку ряду (вибірці), супроводжується великими помилками. Для того, щоб елімінувати надлишкову вагу, надану початковій величині S_0 , Р. Вейд запропонував модифікувати процедуру згладжування таким чином.

Для вихідного моменту часу задається

$$S'_0 = \alpha S_0,$$

$$S'_1 = \alpha x_1 + (1 - \alpha)S'_0 = \alpha x_1 + \alpha(1 - \alpha)S_0,$$

де S_0 – як і раніше, початкова оцінка рівня ряду.

Оскільки коефіцієнти α і $\alpha(1 - \alpha)$ в сумі тепер не дають 1, то варто використовувати множник, який дорівнює одиниці, поділений на суму коефіцієнтів. Таким чином, модифікованим експоненційним середнім для $t=1$ буде

$$S_1 = S'_1 \frac{1}{\alpha + \alpha(1 - \alpha)} = \left[\alpha x_1 + (1 - \alpha)S'_0 \right] \frac{1}{\alpha + \alpha(1 - \alpha)}$$

і взагалі

$$S_t = S'_0 \frac{1}{\sum_{i=0}^t \alpha(1-\alpha)^i} = \left[\alpha x_1 + (1-\alpha)S'_t \right] \frac{1}{\sum_{i=0}^t \alpha(1-\alpha)^i}.$$

Таблиця 12.3 – Зміна ваг в початковий період часу при $\alpha=0,1$ в модифікованій моделі

Ітерація	Вага початкової величини	Вага першого члену ряду	Вага другого члену ряду	Вага третього члену ряду	Вага четвертого члену ряду
1	0,474	0,526			
2	0,299	0,332	0,369		
3	0,212	0,236	0,262	0,291	
4	0,160	0,178	0,198	0,220	0,224

З табл. 12.3 можна помітити, що сутність цього методу полягає в тому, щоб забрати надлишкову вагу від ваги, що надається початковому значенню S_0 , і розподілити її пропорційно по всіх членах ряду. Прогнози, що отримуються за відповідною модифікованою моделлю, базуються більшою мірою на фактичних даних, ніж на попередній оцінці S_0 навіть при малих вибірках. Для того, щоб скоротити час обчислень, доцільно повернутись до звичайного експоненційного згладжування, коли сума коефіцієнтів $\sum_{i=0}^t \alpha(1-\alpha)^i$ наближається до 1. На основі емпіричного аналізу рекомендується здійснювати такий перехід при сумі коефіцієнтів 0,995. При заданому значенні α можна наперед визначити, на якому кроці варто повернутись до звичайної моделі.

12.2.2.4 Вибір постійної згладжування

Вибору величини постійної згладжування варто приділяти особливу увагу. Пошуки повинні бути направлені на відшукування підстав для вибору найкращого значення. Потрібно враховувати умови, при яких ця величина повинна приймати значення, близькі то одному крайньому значенню, то іншому. Неважко помітити, що при $\alpha=0$, $S_t = S_0$ спостерігається випадок абсолютної фільтрації й повної відсутності адаптації, а при $\alpha=1$ приходимо до так званої найвної моделі $\hat{x}_t(t) = S_t = x_t$, відповідно до якої прогноз на будь-який строк дорівнює поточному, фактичному значенню ряду. На практиці ця модель через простоту користується особливою популярністю.

Вище вже відмічалось, що постійна згладжування характеризує швидкість реакції моделі $\hat{x}_t(t) = S_t$ на зміни рівня процесу, але одночасно визначає і здатність системи згладжувати випадкові відхилення. Тому величині α варто надавати те чи інше проміжне значення між 0 і 1 залежно від конкретних властивостей динамічного ряду.

Як задовільний компроміс рекомендується брати її в межах від 0,1 до 0,3. Ця рекомендація не критично повторюється в ряді робіт. Між тим в [95] показано, що найкращі результати отримуються при $\alpha = 0,9$. Однак, як правило, якщо в результаті випробувань виявлено, що найкраще значення константи α близьке до 1, варто перевірити законність вибору моделі даного типу. Часто до великих значень α приводить наявність в досліджуваному ряді яскраво виражених тенденцій або сезонних коливань (висока персистентність процесу, коли показник Хьорста близький до 1 [11,22]). В цьому випадку для отримання ефективних прогнозів необхідна інша модель.

Зрозуміло, що найкраще значення α в загальному випадку повинне залежати від строку прогнозування τ . Для кон'юнктурних прогнозів більшою мірою повинна враховуватись свіжа інформація. При збільшенні періоду випередження τ більш пізня інформація, що відображає останню кон'юнктуру, повинна, скоріш за все, мати дещо меншу вагу, ніж у випадку малих τ . Для того, щоб згладити кон'юнктурні коливання, варто більшою мірою враховувати інформацію за минулі періоди часу. Для проведення подібного аналізу вводять поняття середнього віку даних. Вік поточного спостереження дорівнює 0, вік попереднього спостереження дорівнює 1 і т. д. Середній вік – сума зважених віків даних, використаних для підрахунку згладженої величини. Причому віки мають ті ж ваги, що і відповідна інформація. При експоненційному вирівнюванні вага, надана точці з віком k , дорівнює $\alpha\beta^k$, де $\beta = 1 - \alpha$, і середній вік інформації дорівнює:

$$k = 0 \times \alpha + 1 \times \alpha\beta + 2 \times \alpha\beta^2 + \dots = \alpha \sum_{k=0}^{\infty} k\beta^k = \frac{\beta}{\alpha}.$$

Таким чином, чим менше α , тим більший середній вік (ступінь забуття) інформації. Для кон'юнктурних прогнозів значення α , як правило, треба брати більшим, а для більш довгострокових – меншим. Це положення ілюструє рис. 12.12, на якому відображена залежність стандартної помилки прогнозування, яка, зазвичай, приймається за показник точності, від α . Однак характер залежностей, аналогічних тим, що зображені на рисунку, варто вивчати спеціально в кожному конкретному випадку.

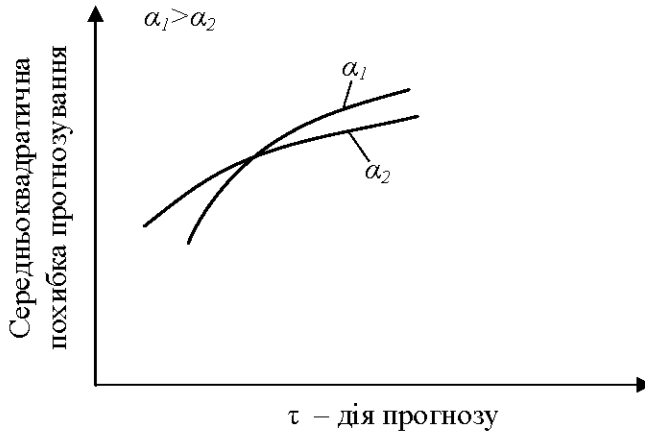


Рисунок 12.12 – Приблизна залежність середньоквадратичної похибки прогнозу від величини постійної згладжування α і періода випередження τ

Теоретичний аналіз проблеми вибору постійної згладжування при використанні найпростішої експоненційної моделі для прогнозування стаціонарного процесу з функцією вигляду $p_k = p_1^k$, де p_1 – коефіцієнт автокореляції при лагу $k = 1$, проведений Д. Р. Коксом і Дж. Д. Кохеном.

Показано, що мінімум середнього квадрата похибки при прогнозуванні такого ряду на 1 крок вперед ($\tau = 1$) буде при

$$\alpha_{opt} = \left\{ \begin{array}{l} \frac{3p_1 - 1}{2p_1}, \frac{1}{3} < p_1 \leq 1; \\ 0, -1 \leq p_1 \leq \frac{1}{3}. \end{array} \right. \quad (12.5)$$

Відповідна дисперсія похибки прогнозу при цьому:

$$D_e = \left\{ \begin{array}{l} \frac{8p_1(1-p_1)}{(1+p_1)^2}, \frac{1}{3} < p_1 \leq 1; \\ \sigma_x^2, -1 \leq p_1 \leq \frac{1}{3} \end{array} \right.$$

Табл. 12.4 показує співвідношення між p_1 , α_{opt} і точністю прогнозування на 1 крок вперед.

Таблиця 12.4 – Співвідношення між $p_1, \alpha_{opt}, D_e / \sigma_x^2$ при прогнозуванні стаціонарного процесу з $p_k = p_1^k$ по моделі експоненційного згладжування $\tau = 1$

p_1	α_{opt}	D_e / σ_x^2	p_1	α_{opt}	D_e / σ_x^2
$\leq 1/3$	0	1	0,7	0,786	0,581
0,4	0,250	0,980	0,8	0,875	0,395
0,5	0,500	0,889	0,9	0,944	0,199
0,6	0,667	0,750	0,95	0,974	0,100

Табл. 12.5 показує, що для даної p_1 величина D_e при $\tau = 1$ слабо залежить від α , так що точність прогнозу в певному околі α_{opt} нечутлива до вибору постійної згладжування.

Результат (12.5) означає, що якщо $p_1 > \frac{1}{3}$, то при відповідному виборі величини α експоненційне середнє у визначеній мірі відображає коливання, пов'язані з сильною автокореляцією.

Таблиця 12.5 – D_e / σ_x^2 як функція від α при $p_1 > \frac{1}{3}$

α	$p_1 = 0,4$	$p_1 = 0,7$	$p_1 = 0,9$	α	$p_1 = 0,4$	$p_1 = 0,7$	$p_1 = 0,9$
1	1,200	0,600	0,200	0,4	0,987	0,647	0,272
0,9	1,136	0,587	0,200	0,3	0,980	0,692	0,318
0,8	1,087	0,581	0,203	0,2	0,980	0,758	0,397
0,7	1,049	0,584	0,211	0,1	0,987	0,853	0,554
0,6	1,020	0,595	0,223	0	1,000	1,000	1,000
0,5	1,000	0,615	0,242	α_{opt}	0,980	0,581	0,200

Певним керівництвом при цьому може слугувати табл. 12.6, яка характеризує дисперсії похибок, отриманих при прогнозуванні стаціонарних процесів з $p_k = p_1^k$, де $p_1 \leq \frac{1}{3}$. З таблиці видно, що при $p_1 < 0$ можна отримати небагато, припускаючи b менше 0,1...0,2. В загальному, очевидно, що якщо $p_1 < 0$, то найпростіша модель експоненційного типу не є хорошим предиктором.

Якщо $\tau > 1$, то суттєво підвищується критична величина, $p_{1крит}$ нижче якої оптимальне значення α дорівнює 0. Цей факт ілюструє табл. 12.7.

Таблиця 12.6 – Дисперсія похибки прогнозу для стаціонарного процесу з $p_k = p_1^k$, де $p_1 \leq \frac{1}{3}, \tau = 1$

α	D_e / σ_x^2				
	$p_1 = 1/3$	$p_1 = 1/10$	$p_1 = 0$	$p_1 = -1/4$	$p_1 = -1/2$
0	1,000	1,000	1,000	1,000	1,000
0,05	1,001	1,020	1,026	1,036	1,043
0,10	1,002	1,041	1,053	1,074	1,089
0,20	1,011	1,087	1,111	1,157	1,190
0,30	1,022	1,139	1,176	1,252	1,307
0,40	1,042	1,197	1,250	1,359	1,442
0,50	1,067	1,263	1,333	1,481	1,600

Таблиця 12.7 – Залежність $p_{1крит}$ від τ

τ	1	2	3
$p_{1крит}$	0,333	0,516	0,821

Рис. 12.13 показує стандартну похибку прогнозування для всіх значень постійної згладжування у випадку стаціонарного процесу з сильною автокореляцією $p_k = 0.9^{|k|}$, тобто автоковаріацією $R_{xx}(k) = \sigma_x^2 0.9^{|k|}$. При цьому пунктирна лінія виділяє геометричне місце точок розв'язку, які мінімізують похибку прогнозування. Звідси можна зробити висновок, що якщо дані сильно корельовані і період випередження τ малий, то згладжувати не варто, а доцільно в якості прогнозу використовувати найбільш пізні спостереження.

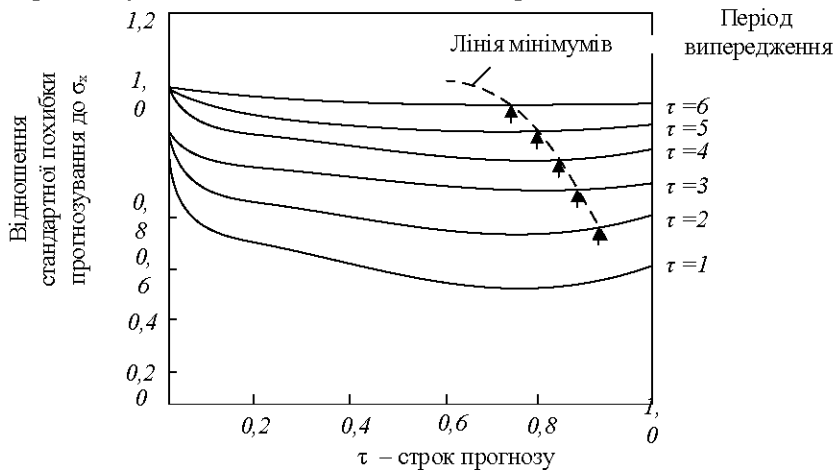


Рисунок 12.13 – Вплив α на точність прогнозування при однократному експоненційному згладжуванні даних з $p_k = 0.9^{|k|}$

12.2.2.5 Формування ансамблів моделей-предикторів

В наш час існує достатньо широкий спектр інструментів інтелектуального аналізу даних для вирішення задач прогнозування: від традиційних методів, подібних розглянутому вище методу експоненційного середнього (ковзного середнього за Р. Брауном), класичних методів статистичного аналізу, розглянутих в розділі 2, до сучасних інтелектуальних методів обробки даних, які використовують дерева рішень, нейро- і нейро-фазі мережі, логістичну регресію і т. д.

Разом з тим різноманітність алгоритмів вилучення знань (Data Mining) говорить про те, що не існує одного універсального методу для вирішення всіх задач. Крім того, використання різних інструментів аналізу й моделювання до одного і того ж набору даних може переслідувати різні цілі: або побудувати спрощену, прозору модель, яка легко інтерпретується, на шкоду точності, або побудувати більш точну, але і більш складну, а значить модель, яка менше інтерпретується.

Таким чином, однією з актуальних задач сучасного підходу до обробки даних, і в тому числі до прогнозування є знаходження компромісу між такими показниками, як точність, складність і здатність до інтерпретування.

Більшість дослідників надають перевагу отриманню більш точних результатів, так як для кінцевих користувачів поняття прозорості суб'єктивне. Точність результатів залежить від якості вихідних даних, предметної області й використовуваного методу аналізу даних.

Отримання більш точних результатів тим більш актуальне, оскільки за останні роки значно зріс інтерес до точності моделей Data Mining, оснований на інтелектуальних методах навчання, за рахунок поєднання зусиль декількох методів і створення ансамблів моделей-предикторів, що дозволяє підвищити якість вирішення аналітичних задач. Під навчанням ансамбля моделей розуміється процедура навчання кінцевого набору базових класифікаторів, результати прогнозування яких згодом об'єднуються і формується прогноз агрегованого класифікатора.

При формуванні ансамбля моделей необхідно вирішити три основні задачі:

- 1) обрати базову модель;
- 2) визначити підхід до використання навчальної множини;
- 3) вибрати метод комбінування результатів.

В силу того, що ансамбль – це агрегована модель, яка складається з окремих базових моделей, то при його формуванні можливі дві альтернативи:

- 1) ансамбль складається з базових моделей одного типу, наприклад, тільки з дерев рішень, тільки з нейронних мереж і т. д.;
- 2) ансамбль складається з моделей різного типу – дерев рішень, нейронних мереж, моделей регресії і т. д.

З іншого боку при побудові ансамбля використовується навчальна множина, для використання якої існує два підходи:

- перевибирання, тобто з вихідної навчальної множини витягується декілька підвбірок, кожна з яких використовується для навчання однієї з моделей ансамбля;
- використання однієї навчальної множини для навчання всіх моделей ансамбля.

В свою чергу, для комбінування результатів, виданих окремими моделями, використовують три основних способи:

- 1) голосування – вибирається той клас, який був виданий простою більшістю моделей ансамбля;
- 2) зважене голосування – для моделей ансамбля встановлюються ваги, з урахуванням яких виноситься результат;
- 3) усереднення (зважене або незважене) – вихід всього ансамбля визначається як просте середнє значення виходів всіх моделей, при зваженому усередненні виходи всіх моделей множаться на відповідні ваги.

Очевидно, що наведене вище ілюструє окремий випадок використання методів багатокритеріального аналізу, зокрема, методу аналізу ієрархій Т. Сааті.

Разом з тим, дослідження в сфері синтезу ансамблів моделей в ІАД (Data Mining) стали проводитись відносно недавно і мають свої особливості і свою термінологію. На даний момент в ІАД розроблено вже багато різних методів й алгоритмів формування ансамблів моделей, серед яких найбільшого поширення набули такі, як бегінг (bagging), бустинг (boosting) і стекінг (stacking).

Алгоритм бегінгу

Головна ідея бегінгу полягає в реалізації паралельного навчання на декількох різних вибірках однакового розміру, отриманих шляхом випадкового відбору прикладів з вихідного набору даних. Алгоритм бегінгу складається з таких кроків. Спочатку формується декілька вибірок шляхом випадкового відбору з вихідної множини даних. Потім на основі кожної вибірки будується класифікатор, і виходи всіх класифікаторів агрегуються з використанням голосування чи простого усереднення. Очевидно, що точність передбачення побудованих за допомогою бегінгу комбінованих предикторів значно вища, ніж точність окремих моделей.

Алгоритм бустингу

Основна ідея бустингу полягає в побудові ланцюжка моделей, при цьому кожна наступна навчається на прикладах, на яких попередня модель припустилась помилки. Порівняно з бегінгом *бустинг є більш складною процедурою* але у багатьох випадках працює ефективніше. Бустинг починає створення ансамблю на основі єдиної вихідної множини, але, на відміну від бегінгу, кожна нова модель будується на основі результатів попередньої, тобто моделі будуються послідовно. Бустинг створює нові моделі таким чином, щоб вони доповнювали раніше побудовані, виконували ту роботу, яку не змогли виконати інші моделі на попередніх кроках. І зрештою, остання відмінність бустингу від бегінгу полягає

в тому, що всім побудованим моделям залежно від їх точності, присвоюються ваги. Бустинг-алгоритм, насправді належить до ітераційних алгоритмів індуктивного моделювання. Він вчиться розпізнавати приклади на границях класів. Кожному запису даних на кожній ітерації алгоритму присвоюється вага. Перший класифікатор вчиться на всіх прикладах з рівними вагами. На кожній наступній ітерації ваги вибираються відповідно до класифікованих прикладів, тобто ваги правильно класифікованих прикладів зменшуються, а неправильно класифікованих – збільшуються. Відповідно, пріоритетними для наступного класифікатора стануть неправильно розпізнані приклади, навчаючись на яких новий класифікатор буде виправляти помилки класифікатора на попередній ітерації.

Алгоритм стекінгу

Стекінг – один із способів створення складових моделей. Даний метод був розроблений нещодавно, тому він менш відомий, ніж бегінг і бустинг. Частково це пов'язано зі складністю теоретичного аналізу, а частково з тим, що загальна концепція використання даного методу поки - що відсутня – основна ідея може бути використана у найрізноманітніших варіантах. На відміну від бегінгу і бустингу стекінг зазвичай застосовується до моделей, побудованих за допомогою різноманітних алгоритмів, що навчаються на однакових даних. Стекінг вводить концепцію метанавчання, тобто намагається навчити кожний класифікатор, використовуючи алгоритм метанавчання, який дозволяє виявити кращу комбінацію виходів базових моделей.