МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

МЕТОДИ МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ І ПРОЦЕСІВ

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня магістр
за освітньою-професійною програмою «Комп'ютерні технології в біології та медицині»
спеціальності 122 Комп'ютерні науки

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2022

Укладачі: *Настенко Є.А.,* д.б.н., к.т.н., професор
Павлов В.А., к.т.н. доцент, доцент
Городецька О.К., к.т.н., доцент, доцент
Корнієнко Г.А., старший викладач

Рецензенти *Степашко В.С*., д.т.н., проф., завідуючий відділом «Інформаційних технологій індуктивного моделювання» Міжнародного науково-навчального центру інформаційних технологій та систем НАН України та МОН України
Шликов В. В., д.т.н., доцент, завідувач кафедри біомедичної інженерії, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

Відповідальний редактор *Зеленский К.Х.*, д.т.н., доцент, доцент кафедри біомедичної кібернетики, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № 2 від 30.09.2022 р.)
за поданням Вченої ради факультету біомедичної інженері
(протокол № 1 від 01.09.2022 р.)

Посібник містить актуальні розділи аналізу та синтезу складних процесів і систем, розглянуто етапи моделювання, викладено основи причинно-наслідкового аналізу, наведено алгоритми та методи індуктивного моделювання, приділено увагу вирішенню нестандартних задач класифікації, аналізу складних систем процесів, застосуванню в моделюванні апарату математичного програмування.

Призначено для здобувачів ступеня магістр, що навчаються за освітньо-професійною програмою «Комп'ютерні науки» спеціальності 122 «Комп'ютерні технології в біології та медицині». Буде корисним для аспірантів та наукових працівників, що спеціалізуються в області комп’ютерного моделювання.

Реєстр. № НП 22/23-131. Обсяг 6,5 авт. арк.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
https://kpi.ua
Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ

БМК – біомедична кібернетика,

ПНА, ПНС, ПНЗ – причинно-наслідковий аналіз, структура, зв’язки,

СПНЗ – статистичні причинно-наслідкові зв’язки,

АКФ – автокореляційна функція,

ВКФ, ККФ – взаємо (чи крос) кореляційна функція,

МНК – метод найменших квадратів,

МІМ – методи індуктивного моделювання,

КАБР – крокові алгоритми багатовимірної регресії,

МГУА – метод групового урахування аргументів,

GMDH – Group Method of Data Handling,

МОС – модель оптимальної структури

БАКСУ - багаторядний алгоритм с комбінаторною селекцією узагальнених змінних,

РІА - Релаксаційні ітераційні алгоритми,

РІАДАПС - РІА на повному дереві структур,

КОМБІ, COMBI - комбінаторний алгоритм методу групового урахування аргументів,

COMBITOS GMDH - комбінаторний алгоритм методу групового урахування аргументів для перетворення простору спостережень (GMDH Combinatorial Algorithm for Transforming the Observational Space)

OSCIA GMDH – Ітераційний алгоритм для перетворення простору спостережень за МГУА (Observation Space Conversion GMDH Iteration Algorithm)

РМНК – рекурентний МНК

СКП – середньо-квадратична похибка

НВСКП - нормована відносна середньоквадратична помилка

АСТ- артеріальний систолічний тиск

АДТ - артеріальний діастолічний тиск

ССС – серцево судинна система

SCP - Set Classification Problem

КМ – класифікація множин

АМП – апарат математичного програмування

ЛП – лінійне програмування

ОДР – область допустимих рішень

y – позначення змінної

$\boldsymbol{y}$ – позначення векторної змінної

РОЗДІЛ 1. ВСТУП У ДИСЦИПЛІНУ

1.1. Особливості завдань біомедичної кібернетики

Біомедична кібернетика є науковим напрямком, що реалізує ідеї, методи, засоби для аналізу, моделювання та оптимізації стану біомедичних об'єктів.

Відмінності біомедичної кібернетика від її попередника - технічної кібернетики, яку називають наукою про управління технічними об'єктами, закладені в особливостях об'єктів дослідження. Специфіка відображається на наступних обставинах і особливостях галузі досліджень:

1. Недостатня вивченість механізмів регуляції станів біологічних об'єктів.

2. Недосконалість технічних засобів вимірювання станів біологічних об'єктів.

3. Етична заборона активного експерименту з людиною.

Окремі винятки, наприклад, клінічні випробування, проводяться в умовах обмежень протоколу, які, як правило, не дозволяють реалізовувати необхідні для побудови якісних моделей варіації вхідних впливів. Тому найчастіше для моделювання використовуються результати пасивного експерименту, моніторингу об'єкту.

Зазначені обставини передбачають акцент досліджень дисципліни в більшій мірі на створенні якісних моделей біологічних систем, ніж на проблемах управління ними, так як без адекватних моделей завдання управління біологічними системами не вирішуються.

Представимо далі дещо узагальнений погляд на моделювання, як на процес проектування і дослідження моделей, а також пов'язаних з цим ідей, механізмів, процедур і проблем.

1.2. Етапи аналізу та синтезу в процедурах моделювання

Узагальнений погляд на моделювання необхідний у зв’язку з тим, щоб:

1. Структурувати послідовність виконання процедур при рішенні завдань моделювання;

2. Виділити загальні механізми і відмінності в існуючому різноманітті завдань моделювання;

3. Розроблювати спеціальні підходи в особливо цікавих, нових для теорії практичних задачах.

Структуризація питань пов'язаних з моделюванням призводить до впорядкування переліку завдань, процедур що стоять між неформальною постановкою проблеми і остаточним математичним та практичним її вирішенням. Є очевидним, що будь-яке завдання моделювання вирішується застосуванням певних процедур аналізу та синтезу. Дані процедури утворюють етапи дослідження, проектування моделей, тут вирішуються [1]:

1. Задачі аналізу об'єкту

2. Задачі синтезу моделі, засновані на результатах аналізу об'єкту

3. Задачі аналізу моделі.

Для динамічних систем (в задачах прогнозу або управління), перед етапом 2 необхідно отримати результати причинно-наслідкового аналізу і тільки після цього можемо приступити до задач синтезу моделі.

Сказане можна підсумувати у вигляді завдань наступних етапів, пов'язаних з моделюванням:

1.Задачі аналізу об'єкту:

- декомпозиція об'єкту,

- прийняття моделестворюючих гіпотез,

- формування складу елементів кожного рівня ієрархії складності об'єкту (змінні об’єкту),

- засобами активного чи пасивного експерименту одержання

матриць об'єкт-властивості спостережень об'єкту.

2.Задачі причинно-наслідкового аналізу [2]:

- побудова статистичних причинно-наслідкових зв'язків системи,
- визначення множини екзогенних змінних системи,
- побудова матриць та графів причинно-наслідкових зв’язків,
- аналіз причинно-наслідкової структури системи,
- побудова причинно-наслідкового фільтру змінних для застосування в задачі синтезу моделей системи.

3. Синтез моделей (класи задач):

- моделювання властивостей даних,
- моделювання систем процесів,
- моделювання систем оптимізації станів біологічних об'єктів,
- моделювання взаємодії систем.

4. Аналіз моделей (властивостей одержаних моделей):

- аналіз адекватності моделей,
- аналіз чутливості моделей,
- аналіз стійкості моделей (для динамічних систем),
- аналіз властивостей динамічних моделей в режимі хаотичного руху,
- аналіз рішень моделей оптимізації.

Наша увага в даному посібнику в основному буде приділена другому етапу (розділ 2), частково третьому етапу в частині розгляду одного з найбільш ефективних підходів до моделювання - методів індуктивного моделювання (розділ 3), та вирішенню нестандартних постановок задач моделювання (розділ 4).

1.3. Контрольні питання

1. Які особливості галузі досліджень відображаються на специфіці завдань біомедичної кібернетики

2. Наведіть суть та охарактеризуйте завдання етапу дослідження об’єкту при моделюванні
3. Наведіть суть та охарактеризуйте завдання етапу причинно-наслідкового аналізу при моделюванні біомедичних об’єктів
4. Наведіть суть та охарактеризуйте завдання третього етапу при моделюванні біомедичних об’єктів
5. Наведіть суть та охарактеризуйте завдання етапу аналізу моделей

РОЗДІЛ 2. ПРИЧИННО-НАСЛІДКОВИЙ АНАЛІЗ

2.1. Моделі прогнозу, автокореляційна та взаємокореляційна функції

Нагадаємо дякі відомості про прогнозування, кореляційні та взаємокореляціні функції. Задачі причинно-наслідкового аналізу зручно розглядати на прикладі моделей прогнозу:

$$y_{n+1} = f(y_n, y_{n-1}, \ldots, y_{n-k}) \tag{2.1}$$

$$y_{n+1} = f(\boldsymbol{y}_{n(0-k)}, \boldsymbol{x}_{1,n(0-k)}, \ldots, \boldsymbol{x}_{m,n(0-k)}) \tag{2.2}$$

де вектор $\boldsymbol{y}_{n(0-k)}=(y_n, y_{n-1}, \ldots, y_{n-k})^T$, вектор $\boldsymbol{x}_{p,n(0-k)}=(x_{p,n}, x_{p,n-1}, \ldots, x_{p,n-k})^T$, n – поточна точка часу, k- максиальне значення лагу враховуємого запізнення.

Графік даних на рис.2.1a) відповідає випадку моделювання по моделі (2.1): прогнозуваня значень ряду «y» по його минулим значенням. Графік даних на рис.2.1 b) – прогнозуваня значень ряду «y» по його минулим значенням та минулим значенням інших зміних "x", покажемо тісний зв'язок між статичною задачею моделювання та задачею моделювання прогнозу на прикладі випадку (рис. 2.1 a):

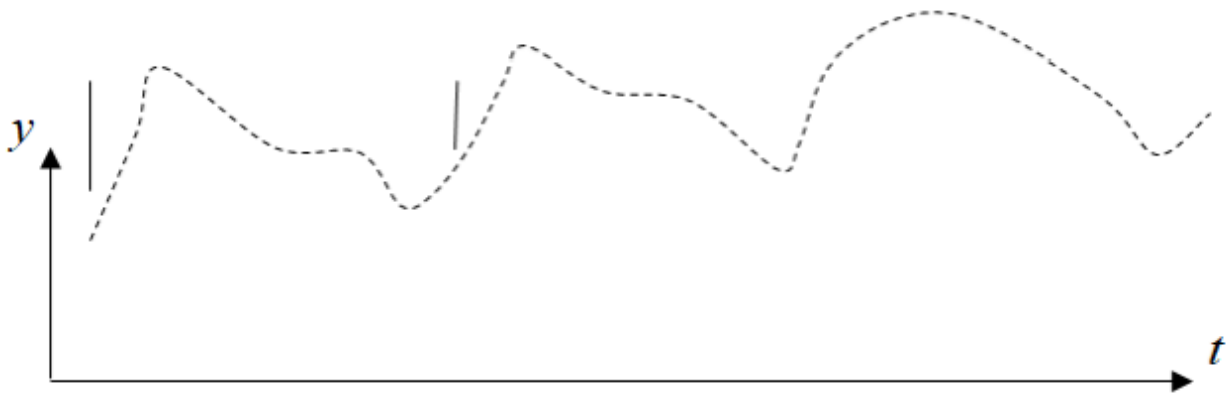


Рисунок 2.1. a) - Графік рядів даних

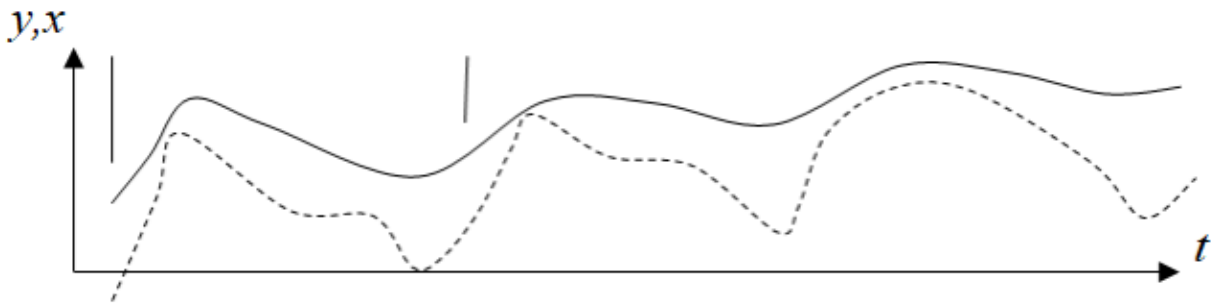


Рисунок 2.1. б) - Графік рядів даних

На рис.2.2 наведено, як споріднені по даним задачі моделювання і моделі прогнозу та статичні моджелі.

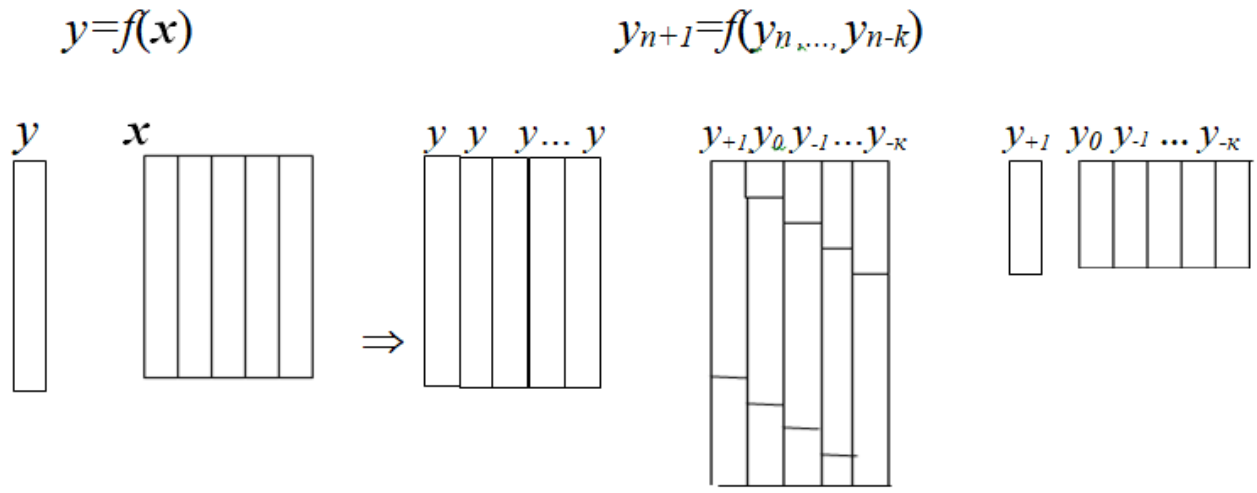


Рисунок 2.2. – Відповідність даних статичного моделювання та прогнозу

Задамося питанням: яким чином визначити до моделювання перспективні запізнювання для введення в модель прогнозу (2.1), що дуже бажано для одержання високого рівня адекватності та хороших прогнозуючих властивостей моделі.

Зрозуміло, що вибір запізнень у моделі прогнозу (2.1) повинен спиратися на автокореляційний ефект, тобто наскільки сильно пов'язані між собою сусідні (з зсувом k) значення ряду, настільки точну і интервально довгострокову модель прогнозу можливо побудувати. Зрозуміло й те , що для

побудови такої моделі корисно дослідити автокорреляційну функцію ряду. Вона безпосередньо вказує на ступінь сили лінійного зв'язку сусідніх значень.

Нижче на рис.2.3 наведено, які ділянки ряду X беруть участь при обчисленні значення автокореляційної функції при зсуві k = 2 [3]:

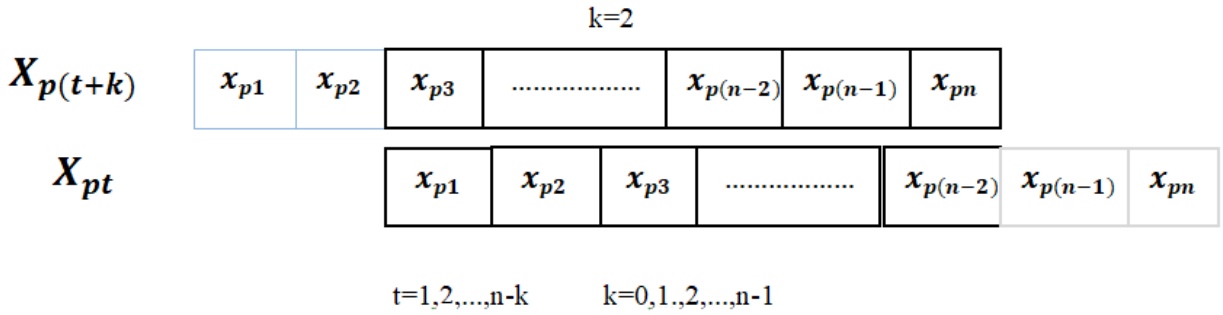


Рисунок 2.3. – Відповідність ділянок ряду X при обрахуванні автокореляційної функції (при k=2)

Формула автокореляційної функції відповідає розрахунку коефіцієнту кореляції ряду з самим з собою, зсунутим на часовий інтервал k.

$$R_k^{X_{p(t+k)}X_{pt}} = \frac{cov(X_{p(t+k)},X_{pt})}{s_p s_p} \tag{2.3}$$

де cov, s_p - виборочні коварації та виборочні оцінки дисперсії відповідних рядів. Индекс зсуву k може і запізнюватися і упереджати, але так як це один и той же ряд X, то функція АКФ парна, симметрична відносно k. Деталізуючи формулу виборочної коваріації одержимо

$$R_k = \frac{\overline{x_{pt} \cdot x_{p(t+k)}} - \bar{x}_{pt} \cdot \bar{x}_{p(t+k)}}{s_p s_p = D} \tag{2.4}$$

Якщо розкриємо формулу через значення часового ряду X , то одержимо (2.5) для розрахунку АКФ через зміщену дисперсію:

$$R_k = \frac{\sum_{i=1}^{n-k} x_i x_{i+k} - \sum_{i=1}^{n-k} x_i \sum_{i=1}^{n-k} x_{i+k} /(n-k)}{\sqrt{\left[\sum_{i=1}^{n-k}\left(x_i - \sum_{i=1}^{n-k} x_i /(n-k)\right)^2\right] \cdot \left[\sum_{i=\kappa+1}^{n}\left(x_i - \sum_{i=1}^{n} x_i /(n-k)\right)^2\right]}} \quad (2.5)$$

або для розрахунку АКФ через незміщену дисперсію:

$$R_k = \frac{\sum_{i=1}^{n-k} x_i x_{i+k} /(n-k) - \sum_{i=1}^{n-k} x_i /(n-k) \sum_{i=1}^{n-k} x_{i+k} /(n-k)}{\sqrt{\left[\frac{1}{(n-k-1)}\sum_{i=1}^{n-k}(x_i - \frac{1}{(n-k)}\sum_{i=1}^{n-k} x_i)^2\right] \cdot \left[\frac{1}{(n-k-1)}\sum_{i=1}^{n-k}(x_i - \frac{1}{(n-k)}\sum_{i=1}^{n-k} x_i)^2\right]}} \quad (2.6)$$

Вигляд автокореляційної функции тісно пов'язаний із виглядом самого ряду. АКФ "білого шуму" (стаціонарний випадок часового ряду), при k > 0, також утворює стаціонарний часовий ряд із середнім значенням 0. Для стаціонарного випадкового ряду АКФ швидко убуває із зростанням k. АКФ, що повільно спадає відповідає наявності виразного тренду у вихідному ряду. Маючи графік автокореляційної функції y можемо визничити значення k зсуву на яких спостерігаємо піки зв’язку значень моделюємого ряду. Включення лагів запізнень ряду із знайденими значеннями зсуву k буде найбіль ефективним для побудови моделі прогнозу ряду y.

Аналогічний аналіз можно навести для обгрунтування застосування кроскорреляційної функції для визначення набору запізнень змінних x для прогнозування змінної y відповідно до моделі (2.2). Маючи на увазі, що кращі значення лагів для змінної y у (2.2) вже визначено, розглянемо додатково взаємо (чи кросс) кореляційні функції – ККФ.

Нехай маємо реалізації $x_{pj}\ j = 1, \ldots, n$ процесів X_p. Виберемо реалізації x_{pj} и x_{qj}, $\mathrm{j} = 1, \ldots, n$ деяких процесів X_p, X_q и побудуємо кросс-

корреляційну функцію з упереджаючим індексом зсуву k для кожного з них.

На рис.2.4 та рис. 2.5 показано формування рядів для розрахунку ККФ $R_{qp}^{k} = R_{X_{qt}X_{p(t+k)}}^{k}$ та $R_{pq}^{k} = R_{X_{pt}X_{q(t+k)}}^{k}$ для k=2

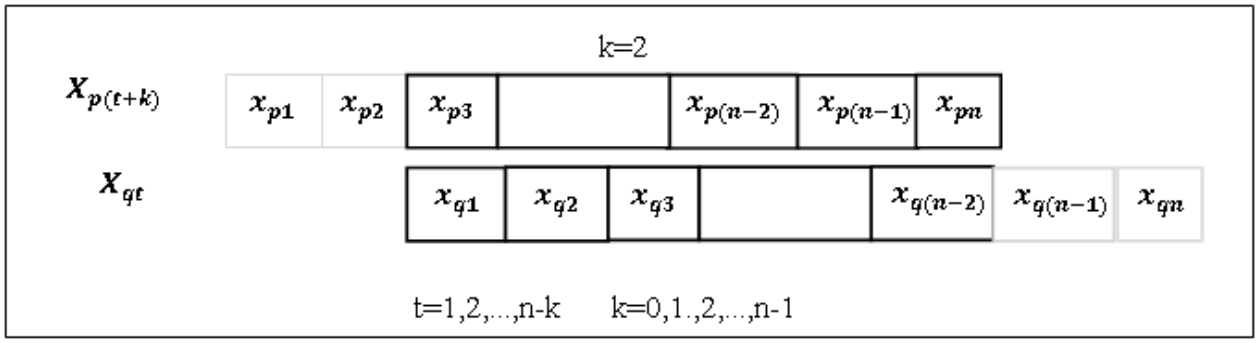


Рисунок 2.4. - ККФ з упереджаючим індексом зсуву k для X_{p}

По елементам рядів, що виділені розраховуємо (рис. 2.4) значення $R_{X_{qt}X_{p(t+k)}}^{k}$

$$R_{qp}^{k} = R_{X_{qt}X_{p(t+k)}}^{k} = \frac{cov(X_{qt},X_{p(t+k)})}{s_q s_p}, \qquad (2.7)$$

де $cov(X_{qt},X_{p(t+k)})$–вибіркова коваріація, s_p , s_q- незміщені вибіркові оцінки дисперсії рядів, що виділені. Деталізуючи формулу виборочної коваріації маємо

$$R_{qp}^{k} = \frac{\overline{x_{qt} \cdot x_{p(t+k)}} - \overline{x}_{qt} \cdot \overline{x}_{p(t+k)}}{s_p s_q} \qquad (2.8)$$

Аналогічно маємо для $R_{pq}^{k} = R_{X_{pt}X_{q(t+k)}}^{k}$, див рис. 2.5

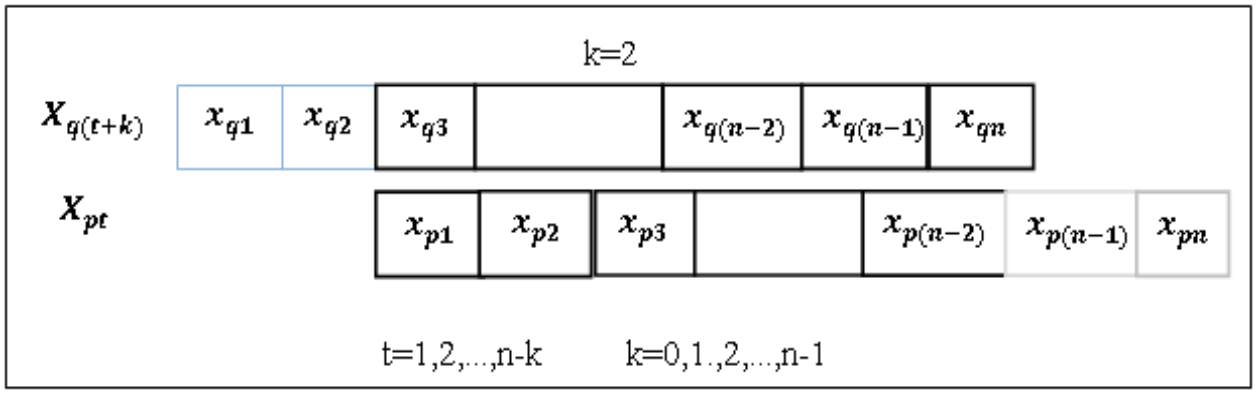


Рисунок 2.5. - ККФ с упереджуючим індексом зсуву k для X_q

По елементам рядів, що виділені розраховуємо значення

$$R^k_{X_{pt}X_{q(t+k)}} = \frac{cov(X_{pt},X_{q(t+k)})}{s_p s_q}, \tag{2.9}$$

де $cov(X_{pt},X_{q(t+k)})$– вибіркова коваріація, s_p , s_q- незміщені вибіркові оцінки дисперсії виділених рядів. Аналогічно (1.8) маємо

$$R^k_{pq} = \frac{\overline{x_{pt}\cdot x_{q(t+k)}} - \overline{x}_{pt}\cdot\overline{x}_{q(t+k)}}{s_p s_q} \tag{2.10}$$

Розрахунок ККФ через вибіркові значення ряду та незміщену оцінку дисперсії:

$$R^k_{X_{qt}X_{p(t+k)}} = \frac{\sum_{t=1}^{n-k} x_{p(t+k)}x_{qt}/(n-k) - \sum_{t=1}^{n-k} x_{p(t+k)}/(n-k)\cdot\sum_{t=1}^{n-k} x_{qt}/(n-k)}{\sqrt{\left[\frac{1}{(n-k-1)}\sum_{t=1}^{n-k}(x_{p(t+k)} - \frac{1}{(n-k)}\sum_{t=1}^{n-k} x_{p(t+k)})^2\right]\cdot\left[\frac{1}{(n-k-1)}\sum_{t=1}^{n-k}(x_{qt} - \frac{1}{(n-k)}\sum_{t=1}^{n-k} x_{qt})^2\right]}\cdot} \tag{2.11}$$

$$R^k_{X_{pt}X_{q(t+k)}} = \frac{\sum_{t=1}^{n-k} x_{q(t+k)}x_{pt}/(n-k) - \sum_{t=1}^{n-k} x_{q(t+k)}/(n-k)\cdot\sum_{t=1}^{n-k} x_{pt}/(n-k)}{\sqrt{\left[\frac{1}{(n-k-1)}\sum_{t=1}^{n-k}(x_{q(t+k)} - \frac{1}{(n-k)}\sum_{t=1}^{n-k} x_{q(t+k)})^2\right]\cdot\left[\frac{1}{(n-k-1)}\sum_{t=1}^{n-k}(x_{pt} - \frac{1}{(n-k)}\sum_{t=1}^{n-k} x_{pt})^2\right]}} \tag{2.12}$$

Розрахунок ККФ через вибіркові значення ряду та зміщену оцінку дисперсії:

$$R^k_{X_{qt}X_{p(t+k)}} = \frac{\sum_{t=1}^{n-k} x_{p(t+k)}x_{qt} - \sum_{t=1}^{n-k} x_{p(t+k)} \cdot \sum_{t=1}^{n-k} x_{qt} / (n-k)}{\sqrt{\left[\sum_{t=1}^{n-k}(x_{p(t+k)} - \sum_{t=1}^{n-k} x_{p(t+k)} / (n-k))^2\right] \cdot \left[\sum_{t=1}^{n-k}(x_{qt} - \sum_{t=1}^{n-k} x_{qt} / (n-k))^2\right]} \cdot} \quad (2.13)$$

$$R^k_{X_{pt}X_{q(t+k)}} = \frac{\sum_{t=1}^{n-k} x_{q(t+k)}x_{pt} - \sum_{t=1}^{n-k} x_{q(t+k)} \cdot \sum_{t=1}^{n-k} x_{pt} / (n-k)}{\sqrt{\left[\sum_{t=1}^{n-k}(x_{q(t+k)} - \sum_{t=1}^{n-k} x_{q(t+k)} / (n-k))^2\right] \cdot \left[\sum_{t=1}^{n-k}(x_{pt} - \sum_{t=1}^{n-k} x_{pt} / (n-k))^2\right]}} \quad (2.14)$$

Для непов’язаних, або слабо пов'язаних випадкових рядів x_i і y_i, $R(k)$ дорівнює нулю або згасає досить швидко. Якщо піки у функції $R(k)$ повторюються через певний час, то взаємний вплив рядів носить періодичний характер. Наявність піків в ВКФ вказує на наявність зв'язку між x_i і y_{i+k} на даному часовому зсуву k і відповідні лаги запізнення доцільно включати у моделі прогнозу виду (2.2).

2.2. Завдання причинно-наслідкового аналізу

Коротко обгрунтуємо завдання ЧНА.

Нехай маємо деяку складну систему взаємо пов’язаних процесів ($x_1, x_2, \ldots, x_m$), для неї відома матриця об’єкт-властивості

$$X = \begin{Bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & ,,, & x_{nm} \end{Bmatrix} \qquad (2.15)$$

Необхідно знайти систему моделей, що описують роботу данної системи процесів.

Перша проблема, з якою стикається дослідник – невизначеність переліку змінних, що підлягають моделюванню. Дійсно, якщо на даному горизонтальному рівні ієрархії складності об'єкту розлядаються для моделювання не одне рівняння прогнозу, а система рівнянь

$$\begin{aligned} x_{1,+1} &= f(\boldsymbol{x}_{1,0(0-k)}, \boldsymbol{x}_{2,0(0-k)}, \dots \boldsymbol{x}_{m,0(0-k)}) \\ x_{2,+1} &= f(\boldsymbol{x}_{1,0(0-k)}, \boldsymbol{x}_{2,0(0-k)}, \dots \boldsymbol{x}_{m,0(0-k)}) \\ x_{m,+1} &= f(\boldsymbol{x}_{1,0(0-k)}, \boldsymbol{x}_{2,0(0-k)}, \dots \boldsymbol{x}_{m,0(0-k)}) \end{aligned} \qquad (2.16)$$

то виникає необхідність визначення, які змінні з списку (2.16) необхідно моделювати, а які ні, тобто необхідно визначити список екзогенних змінних об'єкта, що необхідно виключити з переліку змінних, що включаються до системи (2.16). Цей висновок обумовлює перше завдання причинно-наслідкового аналізу: *визначення множини екзогенних змінних системи процесів.* Очевидно, що множина екзогенних змінних визначається за наступною умовою: якщо визначено, що $x_i \to x_j \ \forall x_j$ то x_i - екзогенна змінна.

Другим завданням ЧНА є *побудова причинно-наслідкового фільтру змінних* «чинників» впливу для кожної змінної «наслідок», що увійшли у список (2.16). Такий фільтр забезпечить наявність у списках моделювання «наслідків» виключно тих змінних, напрямок впливу яких забезпечує коректність побудованих моделей. Для алгоритму моделювання

скорочуються списки вхідних змінних, підвищуються показники адекватності моделі, забезпечується якісна її адекватність.

Третім завданням є *побудова матриць та графів причинно-наслідкових зв'язків.* Доцільність завдання очевидна, бо стає можливим виконання четвертого завдання ЧНА: *аналіз причинно-наслідкової структури системи,* що дозволяє визначити контури, додатні обернені зв'язки та інші особливості структури, що при відповідних умовах можуть привести до особливих режимів роботи системи.

Ну і останнє п'яте, а за виконанням - перше завдання ЧНА: *визначення елементарних статистичних причинно-наслідкових зв'язків системи*. Як очевидно, дане завдання є ключем для виконання інших чотирьох.

2.3. Застосування кроскореляційних функцій для визначення напрямку, матриць та графів причинно-наслідкового зв'язку.

Розглядається задача ЧНА об'єкту який характеризується множинами спостережень, що є реалізації $x_{pj}, j = 1, \ldots, n$ часових рядів (процесів) X_p, $p=1,..,m$ [2].

Ефективність пропонованого далі підходу залежить від ступеню виконання наступних припущень:

1. Часові ряди X_p, що характеризують об'єкт, є взаємопов'язані процеси.

2. Взаємозв'язок процесів X_p в значній мірі характеризується лінійним ефектом.

Розглянемо далі реалізації $x_{pj}, j = 1, \ldots, n$ процесів X_p, $p=1,..,m$ об'єкту дослідження. Виберемо реалізації x_{pj} та x_{qj} . $j = 1, \ldots, n$ деяких процесів X_p, X_q і побудуємо кроскореляційну функцію з упереджуючим зсувом k для кожного з них (див. попередній підрозділ).

Значення ККФ при кожному k характеризують силу лінійного зв'язку між значеннями рядів з даним зсувом, і відповідно ступінь прогностичної можливості відповідної лінійної моделі. Таким чином, відмінність у значеннях $R^k_{X_{qt}X_{p(t+k)}}$ та $R^k_{X_{pt}X_{q(t+k)}}$ можливе застосувати для встановлення напрямку статистичного причинно-наслідкового зв'язку значень ряду при данному k. Інакше кажучі, при $\left|R^k_{X_{qt}X_{p(t+k)}}\right| > \left|R^k_{X_{pt}X_{q(t+k)}}\right|$ лінійний прогноз $X_{p(t+k)}$ по X_{qt} буде більш точний, ніж $X_{q(t+k)}$ по X_{pt} при данному k , тому статистично $X_q \to X_p$. Проте суть причинно-наслідкових відношень є неоднозначною. Вочевидь, що при різних k співвідношення між $R^k_{X_{qt}X_{p(t+k)}}$ та $R^k_{X_{pt}X_{q(t+k)}}$ можуть змінюватися, може змінюватися і напрям впливу процесів. Тому з метою оцінки «взаємовідношень» процесів X_p і X_q і побудови причинно-наслідкових структур доцільно запровадити низку критеріїв, як для окремих значень k, так і інтегральних, для характеристики переважного напряму впливу. Залежно від мети завдання (прогноз, класифікація, аналіз об'єкту) можна запропонувати низку таких критеріїв. Нижче обмежимося найпростішими для визначення напряму ЧНЗ.

Спростимо запис ККФ позначивши $R^k_{X_{qt}X_{p(t+k)}} = R^k_{X_{qt}X_{pt}}$ та $R^k_{X_{pt}X_{q(t+k)}} = R^k_{X_{pt}X_{qt}}$.Тоді для визначення ЧНЗ розглянемо наступні критерії:

$$Cr_1 = Cr_{k=1} = \left|cr_{qp_1}\right| - \left|cr_{pq_1}\right| = \left|R^1_{X_{qt}X_{pt}}\right| - \left|R^1_{X_{pt}X_{qt}}\right| \qquad (2.17)$$

$$Cr_2 = Cr_{n/2} = \left|cr_{qp_{n/2}}\right| - \left|cr_{pq_{n/2}}\right| = \sum_k^{n/2}\left|R^k_{X_{qt}X_{pt}}\right| - \sum_k^{n/2}\left|R^k_{X_{pt}X_{qt}}\right| \qquad (2.18)$$

$$Cr_3 = Cr_n = \left|cr_{qp_n}\right| - \left|cr_{pq_n}\right| = \sum_k^{n-l_{qp}} \left| R^k_{X_{qt}X_{pt}} \right| - \sum_k^{n-l_{qp}} \left| R^k_{X_{pt}X_{qt}} \right| \quad (2.19)$$

$$Cr_4 = Cr_{max} = \left|cr_{qp_{max}}\right| - \left|cr_{pq_{max}}\right| = \left|\max_k R^k_{X_{qt}X_{pt}}\right| - \left|\max_k R^k_{X_{pt}X_{qt}}\right| \quad (2.20)$$

де l_{qp} – один з параметрів процедури, що оптимізується, в залежності від надзавдання аналізу; в звичайному випадку $l_{qp} = l$ - параметр алгоритму, що задається оператором.

З погляду чутливості при визначенні однонаправлених і двонаправлених статистичних причинно-наслідкових зв'язків процедуру встановлення ЧНЗ доцільно параметризувати. Далі в якості параметрів процедури введемо два пороги:

1. Δ_1 – поріг чутливості виизначення напрямку зв’язку ЧНЗ.
2. Δ_2 – поріг значення Cr при котрому правило процедури дійсно.

Для характеристики ЧНЗ системи процесів $X_1, X_2, \ldots, X_m$ введемо матрицю ЧНЗ (рис. 2.6),

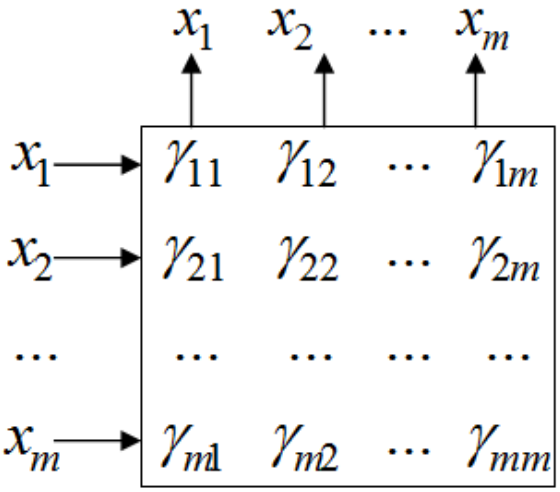


Рисунок 2.6. - Матриця ЧНЗ процесів $X_1, X_2, \ldots, X_m$

для якої елементи $\gamma_{qp} = 1$, якщо $X_q \to X_p$ та $\gamma_{qp} = 0$, якщо $X_q \not\to X_p$

Визначимо далі правило процедури встановлення напрямку зв'язку для всіх варіантів критерію Cr_1, Cr_2, Cr_3, Cr_4 (і відповідних значень елементів матриці ЧНЗ) наступним чином (рис. 2.6):

1.Як при $\max(cr_{pq}, cr_{qp}) \geq \Delta_2$:

1.1. $Cr > \Delta_1$, то $X_q \to X_p$,

1.2. $Cr < -\Delta_1$, то $X_q \leftarrow X_p$,

1.3. Як при $|Cr| < \Delta_1$, то маємо двусторонній зв'язок $X_q \leftarrow X_p$, $X_q \to X_p$.

2.Як при $\max(cr_{pq}, cr_{qp}) < \Delta_2$, якщо:

2.1 $|Cr| < \Delta_1$ - зв'язку між X_p та X_q немає,

2.2 $|Cr| \geq \Delta_1$ - вимагає додаткового дослідження: можливі варіанти

2.2.1 А) зв'язок встановлюється, як у п.1.1,1.2 (пріоритет чутливості) або

2.2.2 Б) працюємо як у п. 2.1 (чутливість ігнорується).

Режим п.2.2 встановлюється, або залежно від результатів додаткових досліджень, або віддається пріорітет режиму 2.2.1, розраховуючи на корекцію структури на етапі моделювання (МГУА спрощує моделі, обриває зайві зв'язки). Внаслідок застосування конкретного критерію при конкретних значеннях порогів отримаємо відповідну даному вибору матрицю ЧНЗ типу рис. 2.6.

Далі наведемо *приклад* матриці та графу ЧНЗ:

Розглянемо спрощений випадок - граф без петіль (тобто не розглядаємо вплив змінних самих на себе). Нехай маємо систему з п'ятьма змінними x_1, x_2, x_3, x_4, x_5, для них згідно розглянутої вище методології визначено напрями причинно-наслідкового зв'язку, що зведено у матрицю ЧНЗ (5 * 5) та побудовано відповідний орієнтований граф ЧНЗ (рис.2.7)

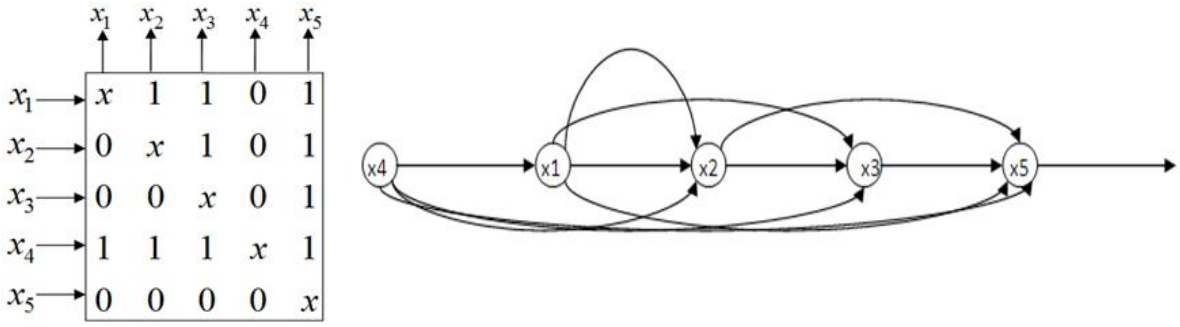


Рисунок 2.7. – Матриця та граф ЧНЗ

Маючи результат у вигляді рис.2.7 можемо вирішити всі завдання ЧНА.

2.4. Причинність по Грейнджеру

Розглянутий вище підхід з точки зору застосування лінійної оцінки ефектів причинно-наслідкового зв'язку має гнучкий характер, так як дозволяє застосувати низку критеріїв для аналізу в залежності від мети дослідження, виду взаємокореляційних функції, суті надзавданя аналізу. Проте історично причинно-наслідковий аналіз представлений статистичною методологією, що має назву «причинність по Грейнджеру», що була заявлена у 1969 році [3]. Основна ідея цього методу полягає у побудові прогностичних моделей, і якщо дані з першого часового ряду x_i допомагають точніше передбачати поведінку другого ряду y_i, то вважається, що змінна x, яка породжує перший ряд – впливає на змінну y, що породжує другий ряд, тобто $x \to y$.

З такого визначення випливає дуже важливій висновок.

Причинність по Грейнджеру може не співпадати з нашим уявленням про фізичне поняття причинності. Це тільки висновок з певного методологічного механізму порівняння адекватності моделей прогнозу, що розглядаються. Далі про це скажемо більш детально.

Метод визначення причинності по Грейнджеру використовується зараз в різних областях, проте одна з головних проблем методу - вдалий/невдалий вибір структури моделей. Нижче наведена одна з реалізацій методу, механізм застосування запропоновано у 2002-2004 роках. Для аналізу причинності будуються індивідуальні моделі, що враховують точки тільки з одного ряду $\{x_n\}_{n=1}^{N}$ чи $\{y_n\}_{n=1}^{N}$, вплив на який оцінюється:

$$x_{n+s} = f_x(x_n, x_{n-l}, ..., x_{n-(D_1-1)l}) \tag{2.21}$$

$$y_{n+s} = f_y(y_n, y_{n-l}, ..., y_{n-(D_2-1)l}) \tag{2.22}$$

З даних моделей оцінюються похибки ε_x та ε_y. Потім будуються спільні моделі, які враховують точки з обох рядів $\{x_n\}_{n=1}^{N}$ та $\{y_n\}_{n=1}^{N}$:

$$x_{n+s} = f_{xy}(x_n, x_{n-l}, ..., x_{n-(D_1-1)l}, y_n, y_{n-l}, ..., y_{n-(D_2-1)l}) \tag{2.23}$$

$$y_{n+s} = f_{yx}(x_n, x_{n-l}, ..., x_{n-(D_1-1)l}, y_n, y_{n-l}, ..., y_{n-(D_2-1)l}) \tag{2.24}$$

З даних моделей оцінюються похибки ε_{xy} та ε_{yx}, тут f_x, f_y, f_{xy}, f_{yx} - поліноми загального вигляду від $D1$, $D2$,($D1+D2$) змінних відповідно, s - дальність прогнозу, $\boldsymbol{l}$ – крок лагу. Для обох моделей параметри розраховуються через МНК.

Коефіцієнти покращення прогнозу, що характеризують причинність по Грейнджеру, виражаються через помилки

$$G_{xy} = \frac{\varepsilon_x^2 - \varepsilon_{xy}^2}{\varepsilon_x^2} \tag{2.25}$$

$$G_{yx} = \frac{\varepsilon_y^2 - \varepsilon_{yx}^2}{\varepsilon_y^2} \tag{2.26}$$

За більшим значенням коефіцієнту покращення прогнозу і визначається напрям впливу.

При $G_{xy} > G_{yx}$ маємо $x \leftarrow y$, та при $G_{xy} < G_{yx}$ маємо $x \rightarrow y$

Але очевидно, що причинність по Грейнджеру має сенс тільки, як ми «вгадали» індивідуальні та взаємні моделі найкращим чином.

Для малих дальностей прогнозу, як правило, лаг береться такий, щоб захопити точку, що лежить через інтервал Z, відповідний нулю автокореляційної функції. Для великих дальностей прогнозу оптимальним вважається лаг, обраний так, щоб захопити точку, що лежить через характерний період s + *l* = T або через два характерних періоду s + 2*l* = T.

Найбільш слабким місцем наведеної методології є:

1.Відсутність раціональних міркувань щодо вибору структури f_x, f_y, f_{xy}, f_{yx}.

2.Відсутність обгрунтувань належної дальності прогнозу *s* та лагу *l*, крім того, лаги *l* у конкурентних моделей можуть бути різні.

3.Сама необхідність визначати 4 структури f_x, f_y, f_{xy}, f_{yx} замість 2-х $f_{x\ по\ y}$ та $f_{y\ по\ x}$ викликає обґрунтовані сумніви.

2.5 Нелінійний статистичний причинно-наслідковий аналіз за МГУА

Дійсно моделі типу

$$x_{n+s} = f_{xy}(y_n, y_{n-l}, \ldots, y_{n-(D_2-1)l}) \tag{2.27}$$

$$y_{n+s} = f_{yx}(x_n, x_{n-l}, ..., x_{n-(D_1-1)l}) \quad (2.28)$$

чи навіть просто

$$x_{n+s} = f_{xy}(y_n, y_{n-1}, ..., y_{n-l}) \quad (2.29)$$

$$y_{n+s} = f_{yx}(x_n, x_{n-1}, ..., x_{n-l}) \quad (2.30)$$

де у (2.29),(2.30) ***l*** - інтервал врахування передісторії, обгрунтовано можливо використувати для визначення напряму впливу *x* та *y*, проте для створення цих моделей необхідно застосовувати математичний апарат, що формує обгрунтовані, оптимальні структури.

Моделі (2.29) та (2.30) дозволяють оцінити *«чистий» взаємний вплив змінних,* що розглядаються. Не врахування ж ефекту впливу у моделях членів типу *xy* чи самих *x, y* для f_{xy}, f_{yx} відповідно, не є обґрунтованими, так як моделі, що розглядаються завідомо є *«фіктивними»* і не враховують ще низку взаємовпливів через *треті змінні*, що достовірно присутні у дійсних прогнозуючих моделях.

Тому, цілком доцільно [2] для встановлення напряму впливу спростити задачу та застосовувати конкуруючі «*фіктивні*» спрощені моделі, що розглядають лише *чистий взаємний вплив* - моделі типу (2.29), (2.30).

Тоді, за умови застосування обгрунтованого методу моделювання, для визначення оптимальних структур спрощених моделей, що моделюють прогностичний эффект лише за рахунок «чистого взаємного впливу», можливо оцінювати напрям впливу просто за більшою точністю (меншою помилкою) прогностичних моделей типу (2.29), 2.30) згідно відповідних величин помилок НВСКП:

$$\delta_{xy} = \sqrt{\frac{\sum_{i=1}^{n}(x_{i+p} - f_{xy}(i))^2}{\sum_{i=1}^{n}(x_{i+p} - \bar{x}_{+p})^2}} \quad , \quad \delta_{yx} = \sqrt{\frac{\sum_{i=1}^{n}(y_{i+p} - f_{yx}(i))^2}{\sum_{i=1}^{n}(y_{i+p} - \vec{y}_{+p})^2}} \qquad (2.31)$$

Напрям впливу визначаємо за умовами:

при $\delta_{xy} > \delta_{yx}$ маємо $x \rightarrow y$, та при $\delta_{xy} < \delta_{yx}$ маємо $x \leftarrow y$ (2.32)

Таким обґрунтованим методом для моделювання (2.29), (2.30) є один з ефективних підходів індуктивного моделювання (з ним ми познайомимося у розділі 3) - метод групового урахування аргументів (МГУА) і у подальшому будемо виходити з цього. Однак причинно-наслідковий аналіз вже не буде методологією встановлення причинності за Грейнджером. Проведений за результатами (2.29), (2.30), (2.31), (2.32) аналіз назвемо *нелінійним причинно-наслідковим аналізом*, а напрям зв'язку між змінними назвемо просто *статистичним причинно-наслідковим зв'язком.*

Так само, як і причинність по Грейнджеру**,** результати *статистичного причинно-наслідкового аналізу можуть* не співпадати з фізичним поняттям причинності. Ми фіксуємо певний напрям *статистичного причинно-наслідкового зв'язку* тільки з висновку деякого показника про статистичну більшу чи меншу точність прогнозу відповідних конкуруючих моделей прогнозу. Більш детально про відміності СЧНЗ та фізичної причинності нижче.

Нагадаємо, що розглядаємо систему змінних (процесів) $x_1, x_2, ..., x_m$ для якої передбачається можливість опису дискретними моделями. Дана система процесів представлена множинами точок $\{x_{11}, x_{21}, ..., x_{n1}\}, \{x_{12}, x_{22}, ..., x_{n2}\}$,…., $\{x_{1m}, x_{2m}, ..., x_{nm}\}$, що формують, як вектора, стовпці відповідної матриці об'єкт-властивості (2.15). Далі розглядаємо наступні пари дискретних моделей із запізнюваннями (розглядаємо спрощені робочі моделі, які будуть

використані для вирішення тільки локального завдання, - встановлення напряму СЧНЗ між парою даних змінних x_i і x_j):

$$\hat{x}_{i,k+s} = f_i(x_{j,k}, x_{j,k-1}, x_{j,k-2}, ..., x_{j,k-l}) \quad (2.33)$$

$$\hat{x}_{j,k+s} = f_j(x_{i,k}, x_{i,k-1}, x_{i,k-2}, ..., x_{i,k-l}) \quad (2.34)$$

Введемо ознаку, по якій надалі визначаємо напрям зв'язку між змінними:

(*) *менш точна* модель вказує в парі змінних x_i , x_j яке направлення зв'язку з двох можливих, найменш ймовірно в даній парі, тобто указує, як *не встановити* напрям причинно-наслідковому зв'язку між ними – $x_i \to x_j$ або $x_j \to x_i$.

За умови максимально об'єктивного характеру отриманих спрощених моделей ознака (*) є *необхідна і достатня умова* для встановлення неіснуючого напряму зв'язку.

З першою ознакою, заснованою на запереченні неіснуючого напряму впливу, зв'язана і *конструктивна ознака:*

(**) точніша модель указує в даній парі змінних x_i , x_j найбільш вірогідний з двох, напрям зв'язку, тобто указує, як *встановити напрям причинно-наслідкового зв'язку* між ними – $x_i \to x_j$ або $x_j \to x_i$.

За умови максимального об'єктивного характеру отриманих моделей ознака (**) є *необхідною* умовою для встановлення поточного напряму причинно-наслідкового зв'язку. Припущення, що дана властивість є *необхідною* умовою відповідного напряму зв'язку є очевидним фактом, що, зокрема, легко показується модельним експериментом.

Приклад.

Наведемо [2] аналіз конкуруючих моделей типу (2.33), (2.34) для доходів та витрат населення України отриманих одним з алгоритмів МГУА. Дані для моделювання використано за період з листопада 1995 по вересень 1999 року.

Ведемо позначення: $x_д$ - доходи населення в Україні (млн. грн.) усереднені за місяць, $x_в$ - витрати і заощадження населення (млн. грн.) усереднені за місяць. Дана пара вибрана з причини очевидності напряму причинно-наслідкового зв'язку в даній парі змінних. Нижче на рис.2.8, наведено графіки процесів доходів $x_д$ (блакитна ломана) та $x_в$ (синя ломана). На рис.2.9 представленно графіки $x_д$ та моделі $x_д$ по $x_в$ (макс. лаг запізнення 10 місяців). На рис. 2.10 представлено графіки Хв та моделі $x_в$ через $x_в$ (макс. лаг запізнення 10 місяців).

За зовнішнім виглядом графіків процесів (рис.2.8) видно, що блакитна крива (доходи) в цілому біжить попереду синьої (витрати), та за часом, характерні повторення витрат відбуваються пізніше (з більшим номером часу), тобто правіше на графіку, чим доходи, що узгоджується із здоровим глуздом.

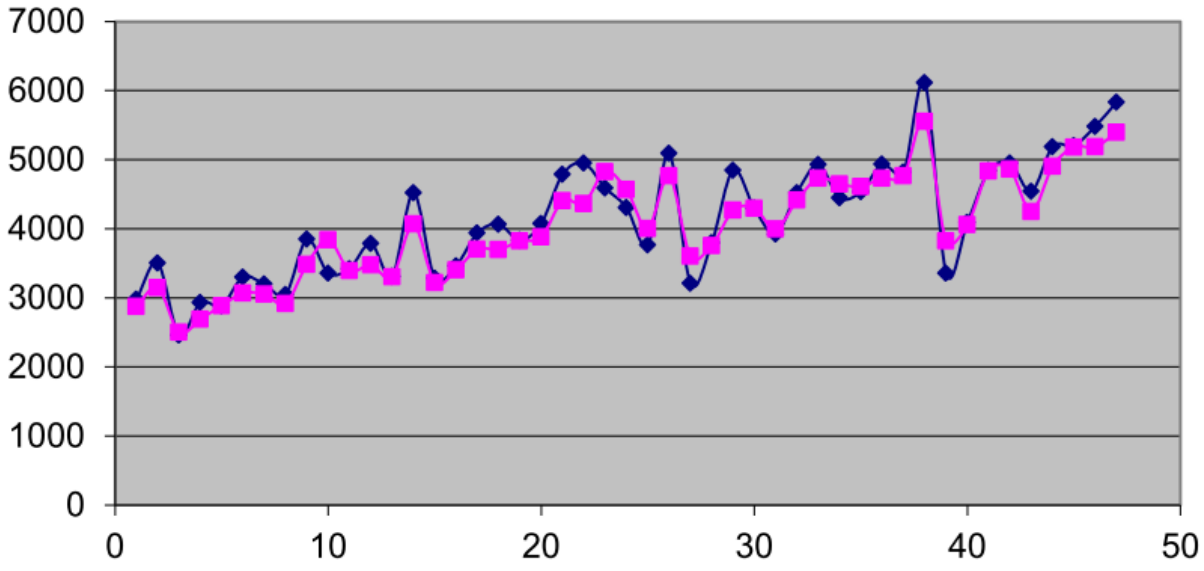


Рисунок 2.8. – Графік даних процесів доходів та витрат населення України

Той же висновок випливає з аналізу точності моделей, графіки яких (блакитним) наближають процеси (синім): рис 2.9 - графік даних $x_д$ та модель $x_д$ по $x_в$, і рис. 2.10 - графік даних $x_в$ та модель $x_в$ по $x_д$. Як видно, більш точна модель (рис. 2.10) наближає витрати $x_в$ по доходам $x_д$. Це слідує і з значень помилок НВСКП - $\delta_{в/д} < \delta_{д/в}$: помилка моделі що моделює доходи по витратам (рис.2.9) $\delta_{в/д} = 0.73$, помилка моделі що моделює витрати по доходам (рис.2.10) $\delta_{д/в} = 0.486$

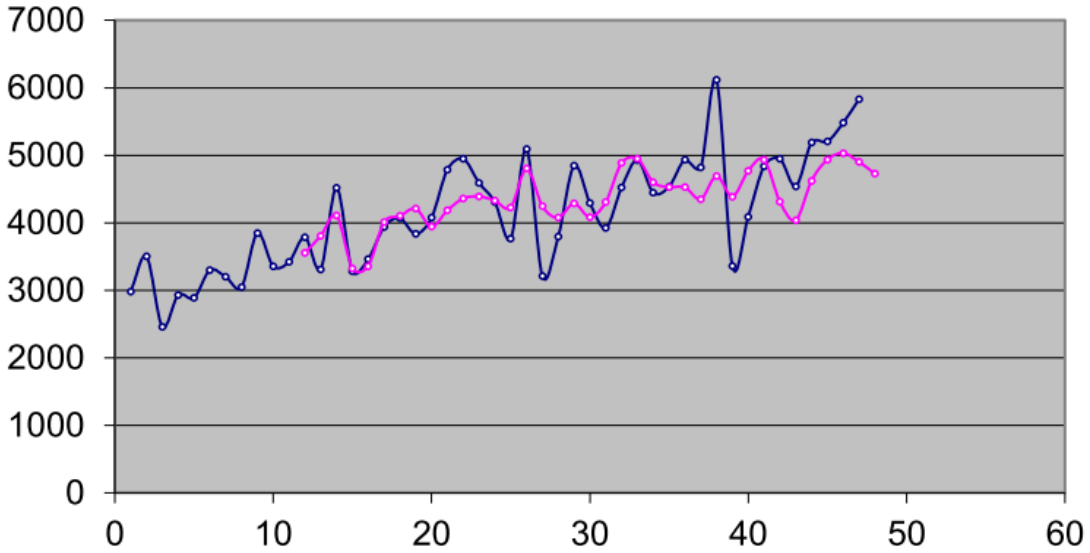


Рисунок 2.9. - Графік даних $x_д$ та модель доходів по витратам.

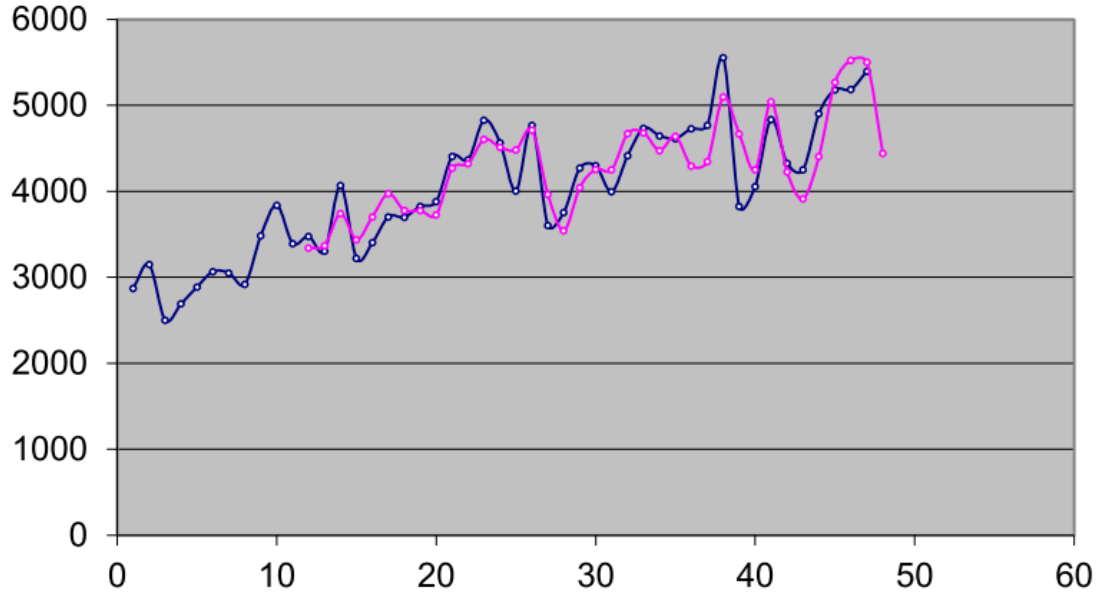


Рисунок 2.10. - Графік даних $x_в$ та модель витрат по доходам.

Відзначимо, що застосування для порівняння спрощених моделей роблять необхідним говорити про ймовірнисний (статистичний) характер певного напряму зв'язку. Про достатність ознаки (**) для визначення напряму зв'язку говорити не можна, як мінімум з двох причин:

а/ простий збіг поведінки розглянутих змінних;

б/ дві дані змінні є два наслідки одного спільного чинника (третьої змінної), при цьому дані наслідки проявляються з різними запізненнями.

Проте, не дивлячись на лише умову *необхідності* ознаки (**) вона є цілком коструктивною для аналізу досить правдоподібного ймовірністного наближення механізму причинно-ніслідкових зв'язків в системі. Дійсно, якщо метою моделювання системи є, наприклад, прогноз, то ні причина а/, ні причина б/ не є перешкодою для включення в розгляд даного зв'язку, якщо він послужить поліпшенню точності прогнозу поведінки системи.

Для забезпечення максимально об'єктивного характеру отримуваних моделей із запізнюваннями (2.33), (2.34) доцільно використовувати метод групового урахування аргументів МГУА, оскільки саме методи самоорганізації дозволяють одночасно визначити і коефіцієнти і форму нелінійного зв'язку між змінними і будують моделі оптимальної складності .

Далі треба зробити істотне зауваження.

Запропонована ознака **(**)** дасть очікуваний результат при порівнянні кращих моделей вигляду (2.33), (2.34). Але перебір всіх можливих структур при всіх можливих значеннях параметрів алгоритму є поки непосильним завданням. Тому доцільно ввести спрощене правило визначення вибору порівнюваних моделей для застосування ознаки (**):

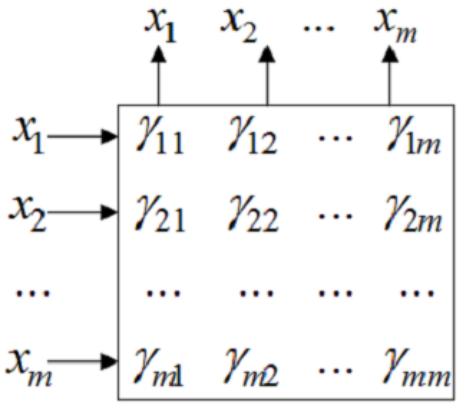
Для деяких, доцільно вибраних параметрів алгоритму, деякого, доцільно вибраного рівня складності (одного і того ж номера ряду селекції, або ж моделей оптимального рівня складності), порівнюємо оцінки точності

отриманої пари моделей. Точніша модель указує в даній парі змінних x_i , x_j найбільш вірогідне з двох, напрям зв'язку, тобто указує, як встановити напрям між ними: $x_i \to x_j$ або $x_j \to x_i$. Поняття доцільності визначиться у кожному окремому випадку об'єкту залежно від розмірності завдання, рівня шумів і інше. На практиці, для значного числа складних об'єктів, розглянута ознака буде не тільки необхідною, але і достатньою для розгляду цілком достовірного наближення причинно-наслідкового аналізу за допомогою отриманих моделей.

2.6. Спрощений та повний порядок встановлення напряму СЧНЗ

2.6.1. Спрощений порядок встановлення напряму СЧНЗ

Структура ЧНЗ системи змінних $x_1, x_2, ..., x_m$ може бути описана і зображена у вигляді матриці та графа. Результат розгляду взаємозв'язків $x_1, x_2, ..., x_m$ в всіх можливих парах x_i , x_j може бути представлений у вигляді матриці на рис. 2.11



$$\gamma_{ij} = \begin{cases} 1, & якщо \quad x_i \to x_j \quad та \\ 0, & якщо \quad x_i \leftarrow x_j \quad чи \quad x_i \not\leftrightarrow x_j \end{cases}$$

Рисунок 2.11. – Матриця причинно-наслідкових зв’язків

Легко визначається і структура графу, що відповідає структурі матриці даної системи. Ми вже розглядали такий приклад на рис.1.7. Одержавши матрицю та відповідний граф, ми зможемо вирішити всі сформульовані нами завдання причинно-наслідкового аналізу. Проте наведена вище методологія причинно-наслідкового аналізу не враховує деякі важливі обставини.

2.6.2. Повний порядок встановлення напряму СЧНЗ

Вище не було враховано вплив на результати аналізу рівня похибки моделей. Наведемо нижче повний порядок встановлення напряму СЧНЗ [2]: Значний рівень похибки (значення нормованої середньоквадратичної погрішності, скажімо, більше 0,9) при моделюванні в парі змінних x_i, x_j дозволяє трактувати пару, яка аналізується, як не зв'язану і ми повинні ігнорувати зв'язок між ними. Якщо ж різниця між оцінками точності (чутливість процедури) отриманих моделей (2.33), (2.34) буде незначна (введемо поріг $\Delta П$), то зв'язок між змінними, як правило, двонаправлений, що може означати в тому числі і їх взаємозалежність через треті змінні.

Ці міркування приводять до більш детальної методології встановлення ЧНЗ. Для формалізації прийняття рішень про напрям ЧНЗ з урахуванням вище вказаних обставин доцільно ввести характерні параметри - δ^* та Δ^*. Нехай в процедурі порівняння конкуруючих моделей $\hat{x}_{i_1}$, $\hat{x}_{i_2}$ змінних x_{i_1}, x_{i_2} похибка моделі $\hat{x}_{i_1}$ по змінній x_{i_2} позначена δ_{i_1}, похибка моделі $\hat{x}_{i_2}$ по змінній x_{i_1} позначена - δ_{i_2}. Тут під δ_i розуміємо НВСКП

$$\delta_i = \sqrt{\frac{\sum_{j=1}^{N_B}(x_{ij} - \hat{x}_{ij})^2}{\sum_{j=1}^{N_B}(x_{ij} - \bar{x}_i)^2}} \qquad (2.35)$$

на тестовій вибірці даних. Тоді

$$\delta_i^I = \min_{i_1,i_2}(\delta_{i_1}, \delta_{i_2}) \tag{2.36}$$

назвемо індикативною похибкою, що вказує індексом i напрям зв'язку між змінними:

$$\begin{cases} i = i_1, \text{ якщо } \min\limits_{i_1,i_2}(\delta_{i_1}, \delta_{i_2}) = \delta_{i_1}^I \text{ та } x_{i_1} \leftarrow x_{i_2} \\ i = i_2, \text{ якщо } \min\limits_{i_1,i_2}(\delta_{i_1}, \delta_{i_2}) = \delta_{i_2}^I \text{ та } x_{i_2} \leftarrow x_{i_1} \end{cases} \tag{2.37}$$

Порогом індикативної похибки $\delta^* = \delta^*_{i_1 i_2}$ назвем граничне найбільше значення індикативної похибки δ_i^I, при якому допускається процедура порівняння конкуруючих моделей. Тобто, при похибці $\delta_i^I > \delta^*$, відповідна модель не може бути застосована для висновку про напрям звязку. Очевидно, чим нижче індикативна похибка δ_i^I, тим більше довіри результату порівняння конкурирующих моделей.

Показником надійності прийняття рішень про чинність в парі змінних x_{i_1} та x_{i_2} назвемо модуль різниці похибок конкурирующих моделей:

$$\Delta_{i_1 i_2} = \left| \delta_{i_1} - \delta_{i_2} \right| \tag{2.38}$$

Мінімальне значення показнику надійності, при якому приймається рішення про орєінтований зв'язок, назвемо порогом показника надійності і позначимо Δ^* ($\Delta^*_{i_1 i_2}$).

Для можливості порівняння результатів у різних парах змінних в одній задачі системного синтезу корисно ввести відносний показник надійності.

Відносним показником надійності η_{ij} приняття рішення про напрям звязку між змінними $x_j \rightarrow x_i$ (стрілка в сторону i) назвемо відношення:

$$\eta_{ij} = \Delta_{ij} / \delta_i^I \tag{2.39}$$

Мінімальна допустиме значення відносного показника надійності (поріг відносного показника) визначимо як:

$$\eta^* = \Delta^* / \delta^* \tag{2.40}$$

Для приведення величин $\eta_{i_1 i_2}$ и $\eta^*_{i_1, i_2}$ в єдиний діапазон, наприклад [0- 100] достатньо змінити формули (2.39) и (2.40) наступним чином:

$$\eta_{i_1 i_2} = \frac{\Delta_{i_1 i_2}}{0{,}01 + \delta^I{}_i}\ , \quad \eta^*_{i_1, i_2} = \frac{\Delta^*{}_{i_1 i_2}}{0{,}01 + \delta^*{}_i} \tag{2.41}$$

Вибір порогів δ^* та Δ^* індивідуальний для кожної задачі. Нижче розглянемо, до чого може привести не оптимальний вибір значень δ^* та Δ^* .

Процедура порівняння конкуруючих моделей, з урахуванням введених показників надійності, виглядає наступним чином:

1. $x_j \in \left(x_j\right)_i^+$ множині змінних, «надійно» направлених в сторону x_i , тобто $x_i \leftarrow x_j$, якщо $\Delta_{ij} \geq \Delta^*$, $\delta^I{}_i = \delta_i \leq \delta^*$, $\delta_i < \delta_j$;

2. $x_j \in \left(x_j\right)_i^-$ множині змінних, «надійно» направлених від x_i , тобто $x_i \rightarrow x_j$, якщо $\Delta_{ij} \geq \Delta^*$, $\delta^I{}_j = \delta_j \leq \delta^*$, $\delta_j < \delta_i$;

3. $x_j \in \left(x_j\right)_i^{\pm}$ множині змінних двустороннього звязку з x_i, тобто $x_i \Leftrightarrow x_j$, якщо $\Delta_{ij} < \Delta^*$ та $\delta^I{}_i \le \delta^*$ (склад цієї множини уточнюємо в подальшому);

4. $x_j \in \left(x_j\right)_i^0$ множині змінних «надійно» не зв'язаних з x_i, якщо $\Delta_{ij} < \Delta^*$ та $\delta^I > \delta^*$;

5. $x_j \in \left(x_j\right)_i^I$ множині змінних, для яких не вдалося вирішити питання про приналежність (Indefinit), ні до однієї з множин, визначених вище, якщо $\Delta_{ij} \ge \Delta^*$ та $\delta^I > \delta^*$.

Скористаємося отриманою інформацією для визначення структури зв'язків у системі. Очевидно, що для моделювання будь-якої змінної x_i з $x_1, x_2, ..., x_m$ необхідно використовувати всі ті змінні, які при аналізі виявилися спрямовані, як $x_i \leftarrow x_j$ (якщо такі маємо), тобто змінні з множин $\left(x_j\right)_i^+$, та $\left(x_j\right)_i^{\pm}$. Якщо множини $\left(x_j\right)_i^+$, $\left(x_j\right)_i^{\pm}$ пусті, то $x_i \in \left(x\right)^{экз}$ тобто належить до множини екзогених змінних даної системи. Таким чином, моделюванню підлягають всі змінні зі списку $x_1, x_2, ..., x_m$, за виключенням множини $\left(x\right)^{экз}$. Зазначений порядок формування списку змінних на вході алгоритму моделювання відповідатиме «жорсткій» фільтрації (максимальному скороченню списку змінних). Використання на вході алгоритму моделювання, крім множин $\left(x_j\right)_i^+$, $\left(x_j\right)_i^{\pm}$, ще множин $\left(x_j\right)_i^I$ та/чи $\left(x_j\right)_i^0$ відповідає різним варіантам «м'якої» фільтрації.

Тепер очевидно, як позначиться на результатах процедури порівняння конкуруючих моделей неоптимальний вибір порогів δ^* та Δ^*. Так

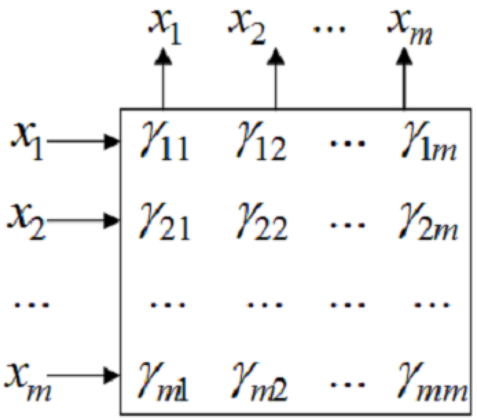
збільшення вимог по надійності (збільшення Δ^*) до процедури порівняння призведе до вимивання змінних з множин $(x_j)_i^-, (x_j)_i^+$, та приєднанню їх до множин $(x_j)_i^0, (x_j)_i^\pm, (x_j)_i^I$. Навпаки, зниження вимог по надійності в процедурі порівняння призведе до виключення змінних з множин $(x_j)_i^0, (x_j)_i^\pm, (x_j)_i^I$ та включення їх до множин $(x_j)_i^-, (x_j)_i^+$. Таким чином, не оптимальність порогів приведе не до помилок у визначенні напрямку зв'язків, а до можливого неповного їх урахування. Наведені міркування у зв'язку з наслідками не оптимального вибору порогів дозволяють зробити висновок, що завищені вимоги по надійності до процедури фільтрації можуть бути компенсовані «м'якістю» варіанту процедури фільтрації.

Серед розглянутих варіантів, як окремий випадок причинно-наслідкового зв'язку, виділяється випадок 3 – множина «одномоментного» двостороннього зв'язку $(x_j)_i^\pm$. Він відповідає причинно-наслідковій взаємодії x_i, x_j, на інтервалі часу, меншому, ніж інтервал дискретності даних Δt ,а в граничному, неперервному випадку, відповідає диференціальному зв'язку змінних на dt.

Виключивши (тимчасово) з розгляду цей окремий випадок, результати дослідження для випадку жорсткої фільтрації можливо уявити квадратною матрицею «*M*» ЧНЗ, де 1 - відповідає наявності зв'язку на вході відповідної змінної, 0 - її відсутності, а незаповнені елементи - невизначеності ситуації з визначенням зв'язку. На рис.2.12 наведено правила формування матриці ЧНЗ

Механізм побудови графу ЧНЗ той же, що раніше. У моделях (2.33), (2.34) використані спільні для обох моделей оптимальної складності, значення параметрів l та s. І якщо значення параметра l (кількість запізнювань) може бути вибрано з умови врахування наявної мінімальної частоти з не нулевою амплитудою в даних (частота тренда), чи на основі розрахунку корреляційної розмірності, то з параметром s - упередженням,

зв’язано певне ускладнення аналізу причинності процесів.



$$\gamma_{ij} = \begin{cases} 1, \text{ якщо } x_j \in (x_j)_i^+ \\ 0, \text{ якщо } x_j \in (x_j)_i^- \cup (x_j)_i^0 \\ \varnothing, \text{ якщо } x_j \in (x_j)_i^I \end{cases}$$

Рисунок 2.12. – Правило формування матриці ЧНЗ

Якщо ми розрахуємо конкуруючі моделі прогнозу для різних упереджень

$s = 1,2,\ldots,p$, то можемо побудувати і відповідні матриці M_s ЧНЗ для різних s .

При припущенні, що матриці M_s для різних s однакові, то співвідношення (типу < >) між похибками δ_i та δ_j конкуруючих моделей при різних s не змінюються, тобто графік залежності похибок конкуруючих моделей від величини упередження буде мати вигляд, як на рис. 2.13, що відповідає напрямку $x_i \to x_j$.

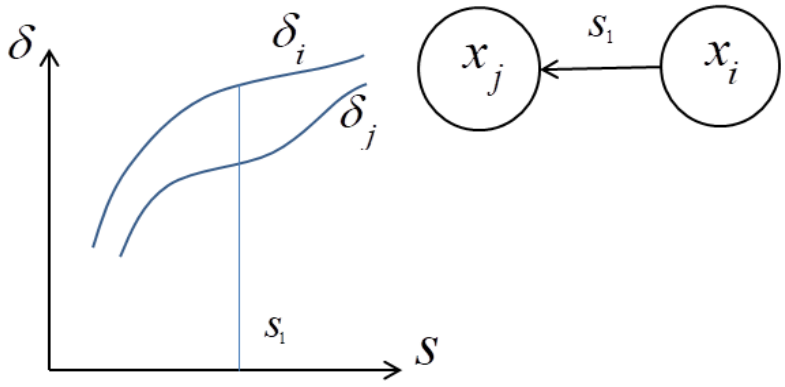


Рисунок 2.13. - Графік залежності похибок конкуруючих моделей від величини упередженням при однакових Ms

У загальному випадку при різних матрицях M_s можемо мати варіанти графіку, подібні до зображених на рис. 2.14, рис. 2.15, коли напрям зв'язку може чередуватися при деяки значеннях s .

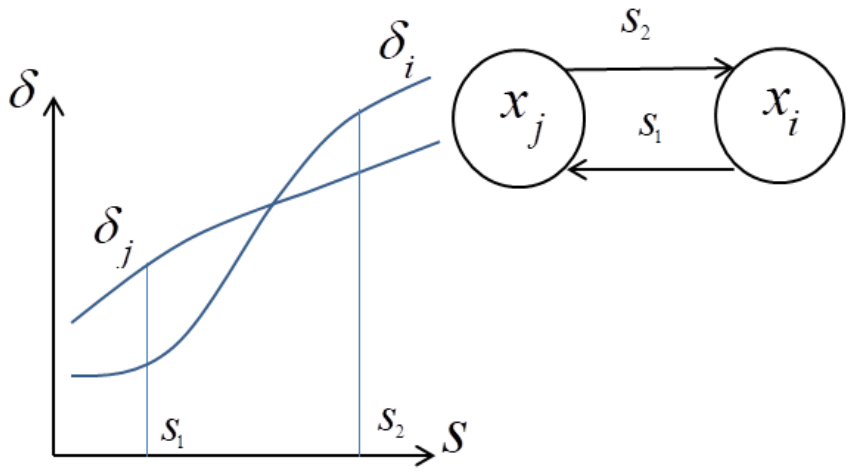


Рисунок 2.14. - Графік залежності похибок конкуруючих моделей від величини упередження при зміні напрямку зв'язку в парі змінних.

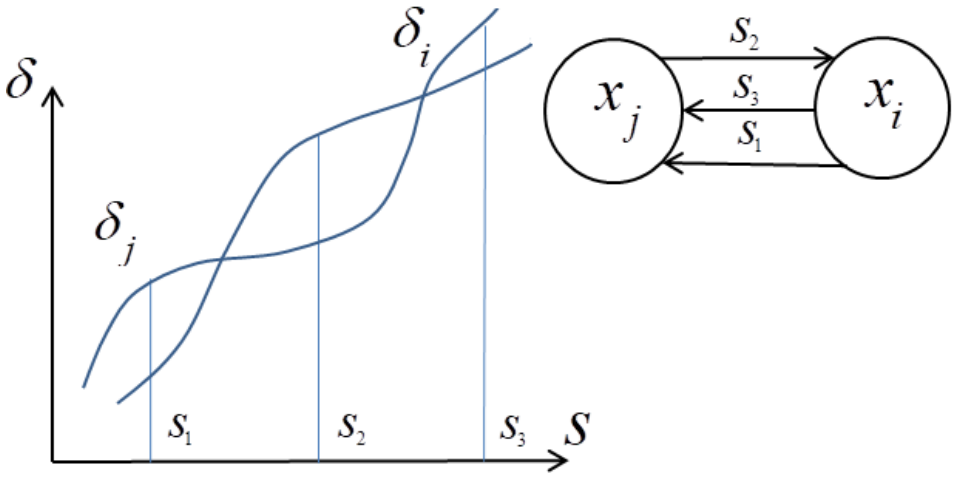


Рисунок 2.15. - Графік залежності похибок конкуруючих моделей від величини упередження при двократній зміні напрямку зв'язку в парі змінних.

Варіанти на рис. 2.14, рис. 2.15 відображають більш складну взаємодію змінних x_i, x_j , на різних частотах, ніж на рис.2.13. Дійсно, амплітуди різних частот що складають процес різні, тому вносять різну частку в загальний

результат при сумуванні. Як наслідок, різні частоти по різному впливають на підсумкове значення моделі і, відповідно, на її помилки при різних упередженнях. Тому доцільно отримувати не одну, а ряд матриць ЧНЗ, що відображають результат взаємодії змінних при кожному окремому упередженні. Такий підхід дозволяє бачити, що на коротких s можемо мати один напрям взаємодії, при інши значеннях s знак взаємодії може змінюватися.

Приклад.

Розглянемо порівняння конкуруючих моделей «доходи населення» та «об'єм виробництва» для економіки України по даним Держкомстата України за період з листопада 1995 р. по серпень 1999 р. Позначимо $x_д$ – доходи населення, $x_о$ – об'єм виробництва. Дані змінні вибрані з міркувань, здавалось би, очевидного результату аналізу, що мав бути одержаний: $x_o \to x_д$. Було розроблено із застосуванням одного з алгоритмів МГУА моделі вигляду (1.33), (1.34):

$$\hat{x}_{д,k+s} = f_д(x_{o,k}, x_{o,k-1}, x_{o,k-2}, ..., x_{o,k-l}) \qquad (2.42)$$

$$\hat{x}_{o,k+s} = f_o(x_{д,k}, x_{д,k-1}, x_{д,k-2}, ..., x_{д,k-l}) \qquad (2.43)$$

для s=1 місяць. Здавалося б, мади одержати $\delta_д > \delta_o$, та очевидний результат $x_o \to x_д$, проте було одержано $\delta_д < \delta_o$, і відповідно, парадоскальний напрямок $x_o \leftarrow x_д$.

Автори посібника були збентежені, проте продовжили дослідження. Перевіряючи одну із гіпотез, одержали моделі при упередженіі s=2, s=3 та s= 4 і виявили, що при s=2 маємо приблизно $\delta_д = \delta_o$, тобто між $x_o, x_д$, при

даному s, існує двонаправлений зв'язок (в якості перехідного режиму) і тільки при $s \geq 3$ одержуємо стійкий характер співвідношення $\delta_д > \delta_o$, тобто напрям $x_o \rightarrow x_д$.

Перебираючі різні обставини та умови функціонування економіки України в період 1995-1999 рр. автори пропонують логічне (пост-фактум аналізу) пояснення одержаного результату: причиною надвеликого впливу "доходів" (доходів у лапках, бо на той період це був практично інфляційний дохід) на рівень виробництва, тобто напрямку $x_o \leftarrow x_д$ при s=1 місяць, став

надвеликий рівень інфляції, що досягав значення 1000% . Ефект помірно стимулюючого (при невеликих значеннях інфляції) та катастрофічно-руйнівного (при надвеликих значеннях) впливу інфляції на економіку давно відомий спеціалістам. Причиною ж встановлення логічного напрямку зв'язку при упередженнях $s \geq 3$ місяці став більш як двомісячний цикл виробницва провідних експортних галузей України. Прикладом є металлургія (цикл виробництва від гірничо-добувного етапу, збагачення руди, виплавлення та прокату сталі, погрузка металу CIF пароплав), що займала на той час не менш 40% об'єму експорту України.

2.7. Обговорення результатів, рекомендації до застосування методології.

Основні результати причинно-наслідкового аналізу концентровані у матрицях ЧНЗ M_s та їх графічних відображеннях - графах ЧНЗ G_s.

Для отримання загальної картини взаємодії у складній системі процесів можно скористатися накладанням матриць суміжності для кожного M_s, отримавши сумарну матрицю M.

За такої матриці можна судити:

- про наявність екзогенних змінних в системі для всіх s,
- про змінні, для яких залишилися не визначеними відносинами в

парах,

- про сумарні взаємодії змінних та їх впливи.

Розберемося з множиною випадку 3 (розділ 1.6.2) класифікації результатів порівняння конкуруючих моделей. З рис. 2.14 та рис. 2.15 та наведеного вище прикладу видно, що даний варіант (рівності похибок) може існувати не стільки, як стійкий двостороній зв'язок, а як перехід знака впливу на протилежний. Тому необхідно, як правило, врахувати дану множину, як підмножину в класі $\left(x_j\right)_i^I$ і зберегти у множині $\left(x_j\right)_i^{\pm}$ тільки ті змінні, для яких умови $\Delta_{ij} < \Delta^*$, $\delta^I{}_i < \delta^*$, (тобто умова нерозрізнюваності в сенсі порогу Δ^*), зберігається для кількох s поспіль. Наявність такої особливості в результатах може свідчити про існування стійкого диференціального зв'язку в системі, що слід врахувати в специфіці синтезу відповідних моделей процесу.

При виборі характерних параметрів δ^* и Δ^* є очевидними наступні рекомендації.

Вимоги надійності прийняття рішення про направлення зв'язку однозначно зрушують поріг Δ^* у бік збільшення, а поріг δ^* у бік зменшення.

З іншого боку, наявність значного числа змінних умножині $\left(x_j\right)_i^I$ свідчить про можливість зменшення порога Δ^*, - в ідеалі множина має бути порожньою. Крім того, доцільно зменшувати поріг Δ^* пропорційно величині передбачуваній кількості змінних m у системі процесів, де беруть участь досліджувані змінні. Проте збільшення рівня шумів в даних робить доцільним відносне збільшення рівня обох порогів δ^* та Δ^*.

Як бачимо, рекомендації, що до значень порогів, є багато в чому суперечливі, тому задача вибору порогів δ^* та Δ^* є оптимізіційною задачою.

Накінець, покажемо можливість використання уточненої версії методології причинно-наслідкового аналізу через кроскореляційні функції.

Врахуємо зв'язок коефіцієнту кореляції з НВСКП:

$$K_{x\hat{x}} = \sqrt{1-\delta_x} \tag{2.44}$$

Можна запропонувати аналіз ККФ – функцій $R^k_{\hat{x}_i x_i}$ та $R^k_{\hat{x}_j x_j}$ побудованих на значеннях кореляції моделей прогнозу $\hat{x}_i(n+k)$ та $\hat{x}_j(n+k)$ на k кроків рядів x_i, x_j по моделям типу (1.33),(1.34), тобто моделюючи $\hat{x}_i(n+k)$ по x_j та його лагам, і моделюючи $\hat{x}_j(n+k)$ по x_i та його лагам. Маємо на увазі що моделі типу (2.33),(2.34) формуються окремо для кожного значення *k*. Тобто, якщо ми при гіпотезі лінійного ЧНЗ між змінними досліджували поведінку кроскореляційних функцій при випередженні на k кроків ряду x_i відносно ряду x_j та навпаки, то вище ми пропонуємо застосувати ту ж методологію при дослідженні та побудові ЧНЗ при гіпотезі нелінійного причинно-наслідкового зв'язку між змінними системи. Тільки у пропонуємому варіанті для дослідження будемо використовувати значення коефіцієнтів кореляції, які будемо розуміти як значення $R^k_{\hat{x}_i x_i}$ та $R^k_{\hat{x}_j x_j}$ при різних *k*.

2.8. Контрольні питання

1. Які методи дослідження причинно-наслідкових зв'язків систем взаємопов'язаних процесів можно назвати, охарактеризуйте кожний підхід.

2. Які завдання доцільно ставити перед причинно-наслідковим аналізом як етапу дослідження в складі процедур моделювання.

3. Які критерії оцінки причинно-наслідкових взаємин досліджуваних процесів можно навести при причинно-наслідковому аналізі за допомогою кроскореляційних функцій.

4. Наведіть основні положення методології дослідження причинності по

Грейнджеру. Охарактеризуйте переваги та недоліки.

5. Аргументуйте, чим пояснюється лише необхідність конструктивної умови для встановлення причинно-наслідкового зв’язку.

6.Наведіть відмінності нелінійного статистичного причинно-наслідкового аналізу за МГУА від аналізу причинності за Грейнджером.

7. Поясніть спрощений порядок встановлення причинно-наслідкового зв’язку.

8. Наведіть повний порядок встановлення причинно-наслідкового зв’язку та обгрунтуйте його.

9. Поясність можливість зміни напряму причинно-наслідкового зв’язку в залежності від величини упередження при аналізі процесів. Наведіть приклад, що ілюструє таку зміну напряму.

10. Наведіть рекомендації до застосування методології причинно-наслідкового аналізу та можливі шляхи її розвитку.

РОЗДІЛ 3. МЕТОДИ ІНДУКТИВНОГО МОДЕЛЮВАННЯ

3.1. Введення в проблеми структурного синтезу

3.1.1. Постановка задачі структурного синтезу

Задано матрицю спостережень розмірністю $X(n,M)$, вектор відгуку $Y(n,1)$, M - кількість аргументів, n - кількість спостережень, передбачається що заданий

А) перелік аргументів $x_1, x_2, ..., x_M$ з надлишком для синтезу моделі, необхідно знайти найкращу модель

$$y = \sum_i^m a_i x_{k_i}\ ,\ m < M\ ,\quad \text{або} \tag{3.1}$$

Б) задається потенційно невичерпний перелік узагальнених вхідних аргументів $z_j = \varphi_j(x),\ j = 1, ..., M, ...,$ де $x = (x_1, x_2, ..., x_t)$ –вектор вхідних змінних. Необхідно знайти найкращу модель

$$y = \sum_i^m a_i \varphi_{k_i}(x) = \sum_i^m a_i z_{k_i} \tag{3.2}$$

Нагадаємо, що обидві постановки відрізняються при розрахунку параметрів

$$\mathbf{a} = (X^T X)^{-1} X^T Y\ ,\ \mathbf{a} = (\boldsymbol{\varphi}^T \boldsymbol{\varphi})^{-1} \boldsymbol{\varphi}^T Y \tag{3.3}$$

тільки необхідністю перерахувати вхідну матрицю

$$X = \begin{Bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & ,,, & x_{nm} \end{Bmatrix} \tag{3.4}$$

у матрицю

$$\boldsymbol{\varphi}(\mathbf{x}) = \begin{Bmatrix} \varphi_1(\mathbf{x}_1) & \varphi_2(\mathbf{x}_1) & \dots & \varphi_M(\mathbf{x}_1) \\ \varphi_1(\mathbf{x}_2) & \varphi_2(\mathbf{x}_2) & \dots & \varphi_M(\mathbf{x}_2) \\ \dots & \dots & \dots & \dots \\ \varphi_1(\mathbf{x}_n) & \varphi_2(\mathbf{x}_n) & ,,, & \varphi_M(\mathbf{x}_n) \end{Bmatrix} \tag{3.5}$$

Ключове питання структурного синтезу - вибір найкращої структури. Термін «найкращості» передбачає відповідність моделі найкращому значенню деякого доцільно заданого критерію якості моделі. Структуру такої моделі називають «структура оптимальної складності».

Зауважимо очевидний факт, що буде використано у подальшому. Як правило, при формуванні нелінійного базису $z_i = \varphi_i(x)$ (наприклад для степеневих функцій) узагальнені аргументи *демонструють певний рівень корельованості* (лінійної залежності).

3.1.2. Моделювання за повним списком аргументів

Спосіб моделювання за повним списком аргументів - найбільш простий метод розв'язання задачі структурного синтезу: модель будується по повному списку аргументів-претендентів (постановка (А)) і виключаються з моделі ті з них, які увійшли в нормовану модель

$$y' = \sum_{i}^{M} a_i' x_i' \tag{3.6}$$

$$x_j' : \quad x'_{ji} = \frac{x_{ji} - \bar{x}_j}{\sigma_{X_j}}, j = 1,...,M, i = 1,...,n \qquad y' : y'_i = \frac{y_i - \bar{y}}{\sigma_Y}, i = 1,...,n \tag{3.7}$$

з рівнем помилки першого роду, меншим ніж (наприклад) 0.05. Підхід можливо рекомендувати застосовувати в умовах, які на жаль практично не спостерігаються в задачах - низьких рівнів шумів в даних і незалежності випадкових вхідних змінних. Крім того, тут не вирішена проблема вибору рівня значущості при якому відбувається виключення елемента структури. А саме це єдиний інструмент, який може компенсувати неідеальні умови структурного синтезу. Зауважимо, що в обумовлених ідеальних обставинах будь-які відомі структурно-параметричні методи будуть давати близькі результати.

3.1.3. Проблеми методів структурного синтезу

Отже наявність зазначених ідеальних умов - ілюзія:

1. Як мінімум, шум обчислень завжди присутній при розрахунку параметрів моделі, і він тим вище, чим гірше обумовленість і вище розмірність інформаційної матриці;

2. Залежність аргументів у практичних завданнях завжди присутня; навіть для даних, породжених генераторами випадкових чисел, формується певний рівень кореляції аргументів - ступінь *лінійної* залежності знаходиться в межах 0.3-0.07. Якщо ж ми звертаємося до постановки Б), то тут факт залежності узагальнених змінних закладено у самій постановці задачі.

Зазначені порушення умов застосування методу призводять до включення фіктивних аргументів у модель замість істинних і зміщеності в оцінках параметрів. І навіть нульова помилка моделювання на навчальній вибірці не вирішує проблему коректного вирішення задачі структурного синтезу. Навіть більше того, можливо навести аргументи, що нульова помилка моделі на реальних даних гарантовано сигналізує про можливе

отримання іншої моделі з кращими властивостями (модель з нульовою помилкою моделює, у тому числі, шум). Причому альтернативна модель, має бути простішою**,** ніж модель з нульовою помилкою. Пропоную вам навести аргументи проти моделей з нульовою помилкою моделювання.

Особливо важливо розглядати стандартну для практики ситуацію, коли моделювання відбувається по корельованим аргументам при наявність шуму в даних. При цьому негативна роль шуму в спотворенні оцінок параметрів багаторазово зростає. Оцінки стають і неспроможними і зміщеними і неефективними. І, чим вище корельованість, тим сильніше ефект спотворення.

Покажемо це, використовуючи геометричну інтерпретацію при отриманні МНК оцінок параметрів моделі Y_x при аргументах X_1 та X_2, як коефіцієнтів проекції вихідного вектору Y_x в площину X_1, X_2. Інтерпретацію покажемо в просторі точок (дивись рис.3.1-3.4).

На рис.3.1 наведено випадок помірної кореляції вхідних векторів. Очевидно, що коли корельованість аргументів зростає, то зменшується кут між векторами X_1 та X_2. При малому куті (рис.3.2) навіть зовсім незначний шум в даних впливає через корельованість, як через збільшувальне скло, на зміщення положення площини аргументів (рис.3.3). І, чим менше кут між векторами, які визначають площину аргументів, тим легше шум може її "повертати" (рис. 3.3)

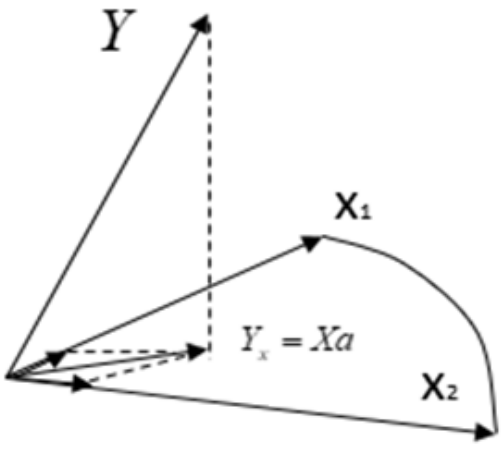


Рисунок 3.1. – Помірна кореляція вхідних векторів

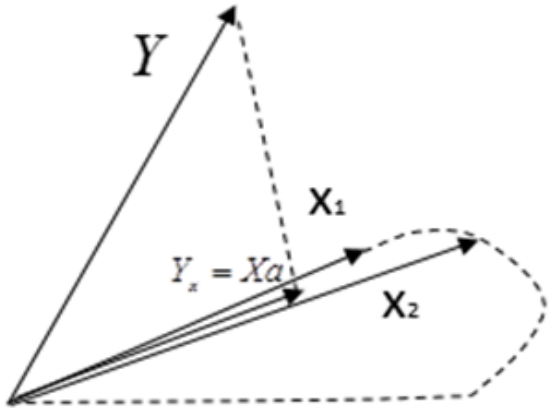


Рисунок 3.2. – Сильна кореляція вхідних векторів

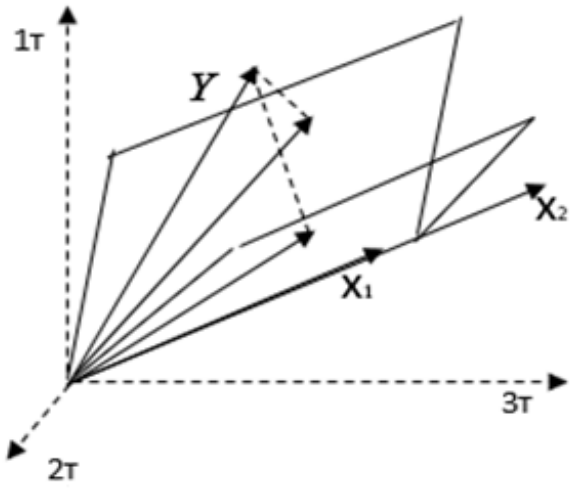


Рисунок 3.3. – Колінеарність вхідних векторів

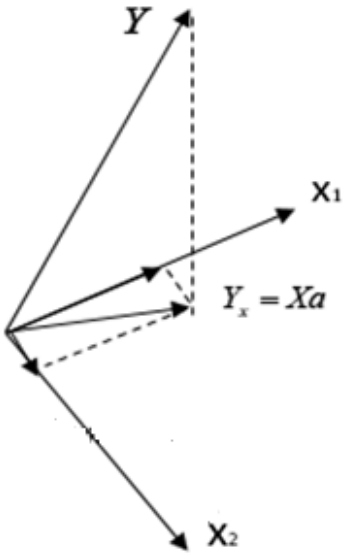


Рисунок 3.4. – Ортогональність вхідних векторів

Рис.3.3 показує випадок близького до колінеарності векторів X_1 та X_2, що дозволяє площині при наявності шуму обертатися на осі векторів і приймати будь-яке положення. Відповідно і проекції (моделі) в ці зсунуті площини будуть, як завгодно відмінні. Саме тому, найкращий варіант проекції – це проекція у площину, визначену ортогональними (незалежними), а ще краще, ортонормованими векторами - рис.3.4. Тут важливо враховувати, що як тільки ми вибрали, якими вихідними векторами **(**X_1 та X_2**)** визначено площину, куди формуємо проекцію - все, модель, по суті, вже зафіксована. Це буде проекція Y_x в дану площину. І вище сказане стосується лише до того, як перезадати цю площину так, щоб:

А) обчислювальний шум мінімально вплинув на процес визначення проекції - необхідно перезадати площину ортогональними векторами і

Б) мінімізувати сам обчислювальний шум, нормувавши вектора аргументів (знизивши діапазон значень елементів інформаційної матриці – покращуємо її обумовленість).

Для реалізації описаних вище процедур існують відповідні інструменти (процедура Грамма-Шмідта [4], оцінки Шелудько [5], тощо). Однак корельованість аргументів і шум у даних відображаються не тільки на оцінках параметрів моделі, але і на структурі моделі. При цьому спостерігаємо не тільки кореляцію вхідних аргументів (як, наприклад, СОЕ і температура, білок в сечі і креатинін в крові), але і з природну залежність узагальнених змінних. В таких умовах перебірні алгоритми стикаються з проблемою множинності рішень**.** Істинні та фіктивні аргументи корелюють, конкурують і претендують на вхід у модель. При цьому фіктивний аргумент, що значно корелює з вихідним, відтісняє істинний аргумент що слабо корелює (входить у модель з малим коефіцієнтом). В результаті, шлях переборного алгоритму до істинної моделі може зациклюватися або взагалі до істинної структури не прийти. Можете спробувати - в цих умовах різні алгоритми дадуть різні фінальні структури. Та й один і той же алгоритм

також дасть вам різні рішення при зміні умов розрахунку:

- варіації вибірки даних,

- зміні у складі узагальнених змінних (при незмінному складі істинних аргументів та значень параметрів моделі),

- зміні у порядку вводу/виключення аргументів в/з моделі,

- зміні в значеннях порогів вводу/виключення аргументів в/з моделі.

У присутності шуму проблеми поглиблюються. Точне рішення можливо знаходити тільки при повному переборі всіх моделей претендентів. Але повний перебір моделей найчастіше неможливий.

Дійсно, розглянемо нижче приклад моделей та узагальнених аргументів, як членів повного поліному Колмогорова – Габора [5]. Кількість ступеневих членів (3.8)

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} a_{ijk} x_i x_j x_k + ..., \qquad (3.8)$$

при m аргументах x та p -тому ступені повного поліному є величина $s = C_{m+p}^{m} - 1$, а кількість моделей які можливо утворити з поліному - $q = 2^s - 1$. Отже, при всього 10-ти вхідних аргументах і 4-му ступені повного поліному, формується опис із $s = C_{m+p}^{m} - 1 = C_{14}^{10} - 1 = \frac{14!}{10! \cdot 4!} - 1 = 1000$ членами, що породжує $q = 2^{1000} - 1$ конкуруючих моделей. То є приблизно 10^{300} моделей, які можливо з нього побудувати. Таке число не тільки зобразити, його уявити неможливо. З наявними обчислювальними ресурсами перебрати і розрахувати та порівняти ці моделі неможливо.

Отже висновок з наведеного аналізу, що до структурного синтезу: кореляція вхідних змінних у присутності шуму породжує необхідність вирішення проблеми множинності моделей і відповідно необхідність перебору величезної кількості моделей претендентів. Причому при організації алгоритмів скороченого перебору успіх залежить від вибору

шляхів такого перебору, серед яких лише деяка меншість (в силу можливої немонотонності критерію селекції на обраному шляху) приведе до моделі істинної структури.

Необхідно запропонувати рішення, що дозволяють відшукати моделі істинної (або фізичної, в постановці відкриття закону) або оптимальної (в постановці задачі прогнозу або класифікації) структури, не вдаючись до процедур повного перебору. Тобто вся множина алгоритмів структурно-параметричного синтезу – це ніщо інше, як множина алгоритмів доцільного скорочення повного перебору структур моделі.

Перед тим як приступити до розгляду можливих рішень проблем структурного синтезу припустимо зараз, що у вас є певний набір інструментів для вирішення даного завдання. Задамося питанням, а чи існує механізм, методологія, що дозволяє сформувати власну думку про алгоритми, що ви маєте застосувати та порівняти, отримати відповідь, який з них найбільшою мірою підходить до вирішення вашого конкретного завдання.

Для відповіді необхідно досконально розібратися у своїй прикладній задачі. Тобто, мати уяву про особливості даних завдання, мати загальні припущення про клас опорних функцій (степеневі, гармонічні ...) найбільш придатних для опису вашого об'єкту і представляти порядок складності істинної моделі. Якщо говорити про моделювання окремого рівня складної (припустимо ієрархічної) системи, то для визначення адекватності методу до вашої конкретної задачі існує ефективний підхід: конструювання модельного прикладу і здійснення модельного експерименту.

3.1.4. Оцінка методу модельним експериментом

Якщо ви розбираєтеся в суті вашої практичної задачі то перевірити метод на адекватність в конкретних умовах досить просто - потрібно створити модельний приклад та провести модельний експеримент [6]: задаємося можливою істинною моделлю, задаємося аргументами x з запасом (справжні та фіктивні), генеруємо вхідні дані і по заданій моделі розраховуємо значення виходу у. Отримані входи і вихід подаємо на вхід

алгоритму, що тестується і аналізуємо, що дав результат моделювання.

Використовуючи принципи та інструменти імітаційного моделювання організуємо імітаційний експеримент і генеруємо деяку множину можливих результатів: варіанти без шуму в даних, з шумом різної дисперсії, розподілу, з аргументами різного ступеня корельованості. Розроблюючи імітаційний експеримент, зручно при фіксації умов моделювання розрізняти три ситуації:

1.Ситуація відповідає постановці завдання у формулюванні А):

Моделі будуються в експерименті на множині $X = \{x_1,..., x_m,..., x_M\}$, тут $X_m = \{x_1,..., x_m\}$- істинні аргументи, інші $\{x_{m+1},..., x_M\}$- фіктивні. При цьому дотримаємося умов:

1а. Вхідна множина змінних *x* - некорельовані змінні і

1б. Шуму в даних немає.

Підготуємо дані для першого етапу експерименту. Задамося можливою істинною моделлю $y^* = \sum_{i=1}^{m} a_i x_i$, і відповідно до умов 1а-1б згенеруємо значення вхідних аргументів $\{x_1,..., x_m,..., x_M\}$. Потім, по заданій моделі розраховуємо значення вихідної змінної y. Всі дані для проведення першого етапу експерименту отримані. Завдання експерименту:

відшукати серед всіх моделей $y = \sum_{x \in X} a_i x_i$, які можна побудувати на X, справжню модель $y^* = \sum_{x \in X_m} a_i x_i = \sum_{i=1}^{m} a_i x_i$.

Очевидно, що в ідеальних умовах 1а-1б будь-який крокový алгоритм, процедура включення аргументу в модель якого заснована на оцінці корисності аргументу за ступенем зв'язку з вихідною змінною, повинен впоратися з відкриттям справжньої структури, так як будь-який аргумент x_i, для котрого $K_{yx_i} \neq 0$, буде істинним. Втрутитися в такий результат може тільки сліпий випадок, тобто перешкодити можуть лише аргументи з множини фіктивних аргументів потужності міри 0, для яких ймовірність

попадання в список аргументів з $K_{yx_i} \neq 0$ дорівнює 0. Покажемо це.

По умовам 1а-1б маємо $K_{x_j x_i} = 0$ для всіх $x_i, x_j \in X$ в силу їх некорельованості і $K_{yx_i} \neq 0$ для всіх $x_i \in \{X_m\}$- істинних (за визначенням, істинний аргумент має зв'язок з виходом).

Припустимо протилежне, що для деякого $x_j \notin \{X_m\}$ виконується $K_{yx_j} \neq 0$. Тоді x_j має бути пов'язаний з виходом y. І при цьому, за визначенням, не є істинним аргументом. Коли таке можливо? - або це випадковість (ідеальний випадковий генератор випадково згенерував такі значення змінної x_j). Тоді теоретично це аргумент з множини аргументів потужності міри 0, ймовірність появи якого дорівнює 0. Або x_j пов'язаний з певним $x_i \in \{X_m\}$, тобто для них $K_{x_j x_i} \neq 0$, що суперечить умові некорельованості всіх $x_i, x_j \in X$. Враховуючи, що звичайний порядок перевірки аргумента-претедента на участь у моделі формується за величиною K_{yx_j}, істинні аргументи послідовно включаться у модель. З огляду на, що ми користуємося в експерименті реальними (а не ідеальними) генераторами випадкових чисел, прийняття рішення про включення аргументу в модель здійснюється за деяким не нульовим рівнем значущості, значення якого повинно відображати якість генератора.

Висновки, які можуть бути зроблені по першому етапу модельного експерименту:

Якщо метод при відсутності шуму в даних і некорельованості вхідних аргументів дає рішення на істинній структурі - це нормально. В іншому випадку, або метод беззастережно не підходить, як той, що не володіє збіжністю до істинної структури, або обрано недостатньо якісний генератор випадкових чисел.

2. Другий етап експерименту проводиться або в постановці А), але при умові корельованості вхідних аргументів або в постановці Б).

У випадку Б), моделі на даному етапі експерименту будуються на множині $\Phi = \{\varphi_1(x),...,\varphi_m(x),\varphi_{m+1}(x),...,\varphi_M(x)\}$, або позначивши $z_i = \varphi_i(x)$, як узагальнені аргументи задачі, на множині $Z = \{z_1,...,z_m,z_{m+1},...,z_M\}$. Тут $Z_m = \{z_1,...,z_m\}$- множина істинних узагальнених аргументів моделі, інші $\{z_{m+1},...,z_M\}$- фіктивні узагальнені аргументи, *x* - вектор вхідних аргументів.

Таким чином умови експерименту припускають:

2а. Корельованість вхідних аргументів *x* (узагальнених аргументів *z*) і

2б. Відсутність шуму в даних.

Згенеруємо вхідні змінні $x = (x_1,...,x_t)$, задамося можливим варіантом істинної моделі в класі постановки А) з кореляцією вхідних аргументів або в класі моделей постановки Б), розрахуємо значення для узагальнених аргументів $z_i = \varphi_i(x)\ i = 1,..,M$ і по заданій моделі отримаємо значення вихідної змінної *y*. Таким чином, отримаємо всі дані для проведення другого етапу експерименту.

Завдання другого етапу експерименту для постановки А):

відшукати серед всіх моделей $y = \sum_{x \in X} a_i x_i$, які можливо побудувати на X, істинну модель $y^* = \sum_{x \in X_m} a_i x_i = \sum_{i=1}^{m} a_i x_i$

Завдання другого етапу для постановки Б):

відшукати серед моделей $y = \sum_{z \in Z} a_i z_i$, які можна побудувати на Z, істинну модель $y^* = \sum_{z \in Z_m} a_i z_i = \sum_{i=1}^{m} a_i z_i$

Висновки, які можуть бути зроблені щодо другого етапу модельного експерименту:

Якщо при умові корельованості вхідних аргументів метод сходиться до істинної структури в більшості практично значимих випадків (на множині варіантів, генерованих імітаційним експериментом), то це добре - значить він

має збіжність до істинної структури, відкидає фіктивні аргументи і відбирає істинні в модель. Ступінь позитивної оцінки методу залежить від ступеня збереження властивості збіжності в міру збільшення ступеня корельованості вхідних аргументів.

3. Третій етап експерименту.

Відрізняється від другого, відмовою від незашумленості даних.

Умови етапу експерименту припускають: різну ступінь корельованості вхідних аргументів x (узагальнених аргументів z) і різну характеристику дисперсію присутнього в даних шуму (обмежимося нормальним розподілом).

Результати, отримані на третьому етапі експерименту можливо характеризувати наступним чином:

Якщо при наявності корельованості аргументів і зашумлення даних метод:

3а. Спрощує модель (в умовах шуму простіші моделі мають менші значення помилок, ніж більш складні і при цьому навіть істинні), забезпечує мінімальну помилку на свіжих даних, це говорить про те, що синтезуються моделі оптимальної складності, і це добре. Для таких моделей будемо мати мінімальні помилки і прогнозу, і апроксимації на нових даних.

3б. Якщо метод, при цьому, породжує модель на підмножині істинних аргументів - це відмінно, так як вирішує відразу 2 завдання - отримує модель оптимальної складності при цьому структура знаходиться на підмножині істинних аргументів.

До якого варіанту структури прагнути - до 3а або 3б, залежить від мети практичного завдання. Якщо стоїть завдання прогнозу або апроксимації в порівняно незмінних умовах, то очевидна корисність варіанта 3а. Якщо стоїть те ж завдання в змінних умовах або отриману модель передбачається застосовувати для розрахунку впливів на досліджуваний процес через деякі вхідні аргументи (тобто передбачається вирішувати завдання управління процесом), то потрібен варіант, максимально наближений до фізичної моделі.

Тобто необхідно прагнути до структури, як в 3б.

3.2. Методи індуктивного моделювання. Огляд підходів

3.2.1. Загальна характеристика підходів індуктивного моделювання

Розглянемо можливі вирішення проблеми вибору структури моделі. Як тепер зрозуміло, найважливіші питання алгоритмів структурного синтезу це вибір критерію якості моделі, найкраще значення якого буде сигналізувати про отримання оптимальної структури і формування шляху перебору моделей на шляху до кінцевого результату. При цьому специфіка різних завдань (апроксимації, прогнозу, класифікації) буде диктувати свої правила побудови критеріїв селекції моделей. Нижче позначимо шляхи вирішення завдання структурного синтезу для багатовимірних регресійних моделей.

Вирішення проблеми проводиться по 2-х напрямках:

1. Відбір найкращих структур здійснюється за рахунок застосування механізму штрафу за складність моделі в явному вигляді.

Механізми враховують доступну інформацію про шуми у вихідних даних і дозволяють отримувати оптимально спрощені (по відношенню до навіть істинної структури) моделі. Отримують оптимальні моделі, наприклад, з точки зору точності прогнозу на тестових даних.

2. Відбір найкращих структур здійснюється за рахунок застосування штрафу за складність моделі в неявному вигляді.

Даний механізм частіше застосовується в найбільш складних умовах, коли про шум не відома інформація, нема доцільних припущень. Цей підхід забезпечує вирішення надзавдання задачі моделювання - мінімізацію помилки моделі на нових вибірках.

Перший шлях частково реалізується в класичних крокових алгоритмах багатовимірної регресії Stepwise, Forward Selection, Backward Elimination і при організації відбору моделей за критеріями Акаіке, Шварца, Меллоуза.

Другий шлях заснований на різних принципах формування нових даних,

свіжих вибірок, і тут здійснюється відбір тих структур, які забезпечують мінімум помилки моделей на цих свіжих даних. Методи, які пропонують свою відповідь на питання, як сформувати такі нові вибірки даних - це

А) метод Jackknife або метод складного ножа, запропонований Maurice Quenouille (1949, 1956) [7],[8] та John Wilder Tukey(1958) [9];

Б) метод Bootstrap запропонований Bradley Efron **(** 1979) [10] та

В) метод групового урахування аргументів МГУА, першу версію якого запропонував Олексій Григорович Івахненко (Україна) у 1968 році [11].

Окремо треба розглядати такі, дуже пов'язані з вище вказаними методами напрямки моделювання, як генетичні, еволюційні алгоритми, нейронні мережі.

3.2.2. Методи індуктивного моделювання з явним штрафом за складність моделей

Наведемо наступні частини алгоритму структурно-параметричного синтезу, які присутні в будь-якому алгоритмі даного призначення:

1. Генератор структур.

Це механізм, який породжує структури, до яких «приміряються» експериментальні дані. Тип генератору визначає базисні функції (степеневі, гармонійні, функції з запізнюваннями, тощо) та шаблон для формування часткового опису (моделі-претендента). Генератор може включати в себе вибір шляху (повного/неповного) перебору структур, які претендують на формування моделі оптимальної складності. Іноді спосіб перебору структур зручно розглядати окремо від генератора.

2.Механізм розрахунку параметрів

У ролі апарату розрахунку параметрів моделі може виступати метод найменших квадратів, метод розрахунку оцінок Шелудько, симплекс методу, тощо…

3. Відбір моделей за критеріями селекції (зовнішні критерії)

3.2.2.1 Крокові алгоритми багатовимірної регресії

Не будемо нагадувати відомі нам алгоритми крокової багатовимірної

регресії Stepwise, Forward Selection, Backward Elimination. Відзначимо тільки, що тут в структурі критерію оцінки корисності аргументу моделі загубився штраф за її складність. Дійсно, критерій оцінки якості аргументу, що вводиться, має вигляд

$$f = \frac{SSR_{\Delta k}}{MSE_k} \tag{3.9}$$

$$MSE_k = \frac{RSS_k}{n-k-2}\ ,\ RSS_k = \sum_i^n (y_i - \hat{y}_i^k)^2\ , \tag{3.10}$$

та $\hat{y}_i^k$ - значення моделі з k аргументами в i - тій точці, а

$$SSR_{\Delta k} = SSR_k - SSR_{k-1},\ SSR_k = \sum_{i=1}^n (\hat{y}_i^k - \bar{y})^2 \tag{3.11}$$

Чим вище значення міри f, тим вище якість k -того аргументу. Але в чисельнику міри знаходиться коефіцієнт ($n-k-2$), який зі збільшенням k (складності моделі) зменшує значення показника якості аргументу, що має зміст штрафу за складність моделі.

Однак вплив даного інструменту на структуру моделі в даному алгоритмі невеликий і навіть мізерний при $n >> k$. Основний вплив на структуру має вибір величин порогів рівнів значущості ($\alpha_{вкл}$ та $\alpha_{искл}$) або порогів значення критерію Фішера ($F^{вкл}$ та $F^{викл}$) на включення і виключення аргументу в / з моделі. При цьому основною проблемою при визначенні структури стає саме вибір значень цих порогів, який в класичній версії КАБР має зробити оператора моделювання. Застосування принципів самоорганізації дозволяє коректно вирішити дану проблему (ми покажемо це при розгляді алгоритмів МГУА).

3.2.2.2 Вибір структури по критерію Малоуза

Нехай відома дисперсія шум у в даних $\hat{\sigma}^2$. Можна показати, що найбільш відповіднім критерієм для вибору структури моделі прогнозу оптимальної складності в цих умовах є критерій Малоуза. Він дає структуру моделі, що відповідає незміщеній оцінці прогнозу процесу. Даний критерій відомий у двох формах:

$$C_p = RSS(s) + 2\hat{\sigma}^2 s \quad \text{та} \quad C_s = \frac{1}{\hat{\sigma}^2} RSS(s) + 2s - n \tag{3.12}$$

де *RSS* - квадрат норми нев'язки виходу ***y***, $\hat{\sigma}^2$ - дисперсія шуму, S - складність моделі, n – кількість точок.

3.2.2.3 Вибір структури по критерію Акаіке

Нехай відомий розподіл шуму f у вихідних даних. Це означає, що задана імовірнісна функція втрат $F(x, y)$, що задовольняє принципу найбільшої правдоподібності

$$\ln F(x_i, y_i(a(s))) = \ln(\prod_i^n f(x_i, y_i(a(s))) \tag{3.13}$$

де $F(x, y)$ - щільність імовірнісного розподілу на множині X×Y, відповідно якій одержано навчальну вибірку (x_i, y_i) $i = 1,...,n$. Для знаходження параметрів $\hat{a}(s)$ використовують метод найбільшої правдоподібності. Знайдемо такі значення $\hat{a}(s)$, що доставляють максимум функції правдоподібності $\ln F(x_i, y_i(\hat{a}(s))) = \ln(\prod_i^n f(x_i, y_i(\hat{a}(s)))$ для варіанту структури S у відомих точках (x_i, y_i). Тоді для пошуку оптимальної структури S застосовується інформаційний критерій Акаіке

$$AIC(s) = 2\ln F(x_i, y_i(\hat{a}(s))) + 2s \tag{3.14}$$

де $F(x_i, y_i(\hat{a}(s)))$ - максимальне значення функції правдоподібності

моделі. З урахуванням того, що у випадку лінійної регресії шум передбачається нормальним, а процедура пошуку методом найбільшої правдоподібності зводиться до МНК, інформаційний критерій Акаіке може бути записано у вигляді

$$AIC(s) = n\ln(RSS/n) + 2s \tag{3.15}$$

На практиці він часто застосовується в спрощеному вигляді

$$AIC(s) = RSS(s) + 2s \tag{3.16}$$

Цей варіант формули називають критерієм Акаіке-Малоуза. Даний критерій істотно обмежує зростання складності моделі наявністю адитивного члена 2s. Однак проблема застосування полягає в тому, що в практичних завданнях функція розподілу шуму і його дисперсія невідомі. А що тоді робити? Використовують менш обґрунтовані, проте теж непогано працюючі критерії.

3.2.2.4. Байєсівський інформаційний критерій (критерій Шварца):

$$BIC(s) = \ln(MSE) + \frac{(s+1)\ln(n)}{n} \tag{3.17}$$

Також є популярним критерій фінальної помилки передбачення Акаіке, що застосовується при невідомому характеру шуму і корегуючий залишкову суму квадратів помилки

$$FPE(s) = \frac{(n+s)}{(n-s)} RSS(s) \tag{3.18}$$

3.2.3. Методи індуктивного моделювання з неявним штрафом за складність моделей

Дані методи використовують критерії штрафу за складність моделі в неявному вигляді та з породженням нових вибірок. Зміст критеріїв: вибирається та структура, яка забезпечила досягнення мінімуму помилки на нових вибірках.

3.2.3.1. Характеристика методу Bootstrap

Метод Bootstrap передбачає, що раз конкретні *n* точок з'явилися у нас у навчальній вибірці, то у них і в інших навчальних вибірках рівна ймовірність появи. Дана гіпотеза лежить в основі алгоритму породження нових подібних вибірок розмірністю *n*. Для цього використовуються алгоритми імітаційного моделювання нових вибірок за допомогою рівномірного імовірнісного розподілу. Ключевий момент: якщо деяка точка реалізувалася в поточну нову вибірку, вона повертається знову у множину генерації. Таким чином отримуємо вибірки розміром *n* з деякою можливою кількістю повторень окремих точок.

3.2.3.2. Характеристика методу Jackknife

Метод використовує критерій який називають "ковзний контроль", або "усереднений критерій регулярності", або критерій "Джекнайф" - складаний ніж. Метод використовується при вкрай малій кількості точок. Коли точок просто замало використовують розбиття вибірки по МГУА. Критерій має вигляд:

$$JN(s) = \frac{1}{n}\sum_{i-1}^{n} MSE_i^{n-1}(s) \qquad (3.19)$$

де $MSE_i^{n-1}(s)$ - значення критерію MSEi при синтезі моделі на (n -1) точці. Мається на увазі, що критерій розраховується при вилученні з вибірки деякої i-тої точки при данному s. Коли s визначено - перераховуємо параметри на повній вибірці.

3.2.3.3.Характеристика методу групового урахування аргументів

Метод МГУА [5] застосовує розрахунки критеріїв на підвибірках загальної робочої вибірки. Принцип штрафу (в неявному вигляді) досягається за рахунок вибору структури при досягнення екстремуму на вибірці, що не була застосована при розрахунку параметрів. По значенню критерію на одній частині, навчальній (називають внутрішнім критерієм), здійснюється оцінювання параметрів, по іншій, перевірочній (називають зовнішнім критерієм, або критерієм селекції), визначається здатність структури моделей для прогнозування. Найбільш часто для оцінки структури застосовується критерій регулярності:

$$\Delta_B^A(s) = \left\| y_B - X_B \hat{\theta}_{A_s} \right\|^2 \qquad (3.20)$$

де A і B - навчальна і перевірочна частини вибірки, $\hat{\theta}_{A_s}$ - розраховані за МНК параметри моделі складності s на підвибірці A. Вид норми у (3.20) буде наведено нижче при детальному розгляді алгоритмів МГУА.

Далі більш детально розглянемо один з основних підходів до вирішення проблем структурного синтезу - метод групового урахування аргументів МГУА. Таку нашу увагу методу виправдано тим, що МГУА не просто метод, а цілий напрямок у моделюванні, що породив десятки алгоритмів у яких спільними є лише загальні принципи, що лежать в основі концепції МГУА. Підхід успішно застосовується в різних типах завдань моделювання (апроксимації, прогнозу, класифікації), завданнях причинно-наслідкового і системного аналізу в різних прикладних областях. Ідеї методу знайшли відображення в генетичних алгоритмах, підходах до оптимізації структури нейронних мереж. Однак головна причина нашої уваги до підходу є те, що метод безперервно розвивається і стає ключем до вирішення нових класів задач, що виникають перед дослідниками, основою для множини нових унікальних алгоритмів вирішення задач моделювання.

3.3.Метод групового урахування аргументів МГУА.Теорія.

3.3.1. Поняття зовнішнього критерію. Критерій регулярності

Ідея зовнішніх критеріїв в МГУА з'явилася, як ідея налаштування структури моделі на новій, свіжій інформації, що не використовувалася для оцінки параметрів (для оцінки параметрів використовуються внутрішні критерії). Ця ідея дозволяє розумно вирішити проблему множинності пропонованих алгоритмами моделей, проблему, як знайти серед цих моделей істинну, або хоча б ту, яка б вела себе, як істинна, щодо тих властивостей, які нас цікавлять перш за все. Образно кажучи, ідею зовнішніх критеріїв можна ілюструвати таким сюжетом: нехай для рішення деякої проблеми суб'єкту необхідно прийняти одночасно двоє несумісних ліків. Зрозуміло, що побічний ефект для вирішення проблеми буде надзвичайно неприємним. Але якщо проблему вдалося переформулювати таким чином, щоб розділити суб'єкта прийняття ліків - ліки прийняли двоє окремих суб'єкта, то проблема буде вирішена. Саме поділ області визначення зовнішнього і внутрішнього критеріїв дозволяє ефективно вирішувати завдання оптимізації структури.

3.3.1.1. Формальне визначення зовнішнього критерію

Під зовнішнім критерієм розуміється міра найбільш важливої для визначення структури моделі властивості (точність апроксимації, прогнозу, дискримінації, гладкості ...) обчислена за інформацією, що не використовувалася при оцінці параметрів моделі. Один з найбільш поширених зовнішніх критеріїв МГУА - критерій регулярності Δ_B^A [5] .

3.3.1.2. Критерій регулярності

Нехай повна вибірка даних позначена W (*learning*). Поділимо її на 2 підвибірки A і B такі, що $W = A + B$, $A \cup B = \varnothing$, Тоді A називають навчальною вибіркою (*training*) на якій оцінюємо параметри моделі. Для неї нормована середньоквадратична помилка $NSSR$ дорівнює Δ_A^A :

$$\Delta_A^A = \sqrt{\frac{\sum_{N_A}(y_A - X_A\hat{\theta}_A)^2}{\sum_{N_A}(y_A - \bar{y})}} \tag{3.21}$$

де $\hat{y} = X\hat{\theta}_A$ - модель, параметри якої $\hat{\theta}$ знайдені на A. Вибірку B - називають перевірочної вибіркою (*testing*) на якій відбираємо структури. Нормована середньоквадратична помилка *NSSR* тієї ж моделі на B дорівнює

$$\Delta_B^A = \sqrt{\frac{\sum_{N_B}(y_B - X_B\hat{\theta}_A)^2}{\sum_{N_B}(y_B - \bar{y})}} \tag{3.22}$$

яку і називають критерієм регулярності . Зміст критерію: якщо модель без налаштування параметрів, а тільки за рахунок структури (адже параметри оцінюємо на A) непогано вгадує прогноз або апроксимацію на свіжих точках (на B), то напевно і на деяких інших, майбутніх точках (майбутня вибірка C) слід очікувати непоганих результатів.

Дана ідея і відповідний підхід вперше запропонований в Україні в школі академіка Івахненко Олексія Григоровича в 1968 р. і отримав назву методу групового урахування аргументів (МГУА) [11]. Але ідея МГУА набагато ширше критерію регулярності - це підхід про необхідність вибору зовнішнього критерію для відбору структур.

Наприклад, в задачах побудови обтічників зовнішнім критерієм служить не точність на B, а критерій гладкості моделі. А при прогнозуванні трендів цін на біржі зовнішнім критерієм служить кризовий критерій (кількість збігів зміни знаків похідної у моделі та об'єкту). Різноманітність зовнішніх критеріїв дало можливість ефективного застосування МГУА практично у всіх типах завдань моделювання: прогнозу, апроксимації, відкриття законів, класифікації, ЧНА в різних областях досліджень.

3.3.2. Загальна схема МГУА

Загальну ідею алгоритмів МГУА можно демонструвати на спрощеній

схемі багаторядного алгоритму (рис.3.5):

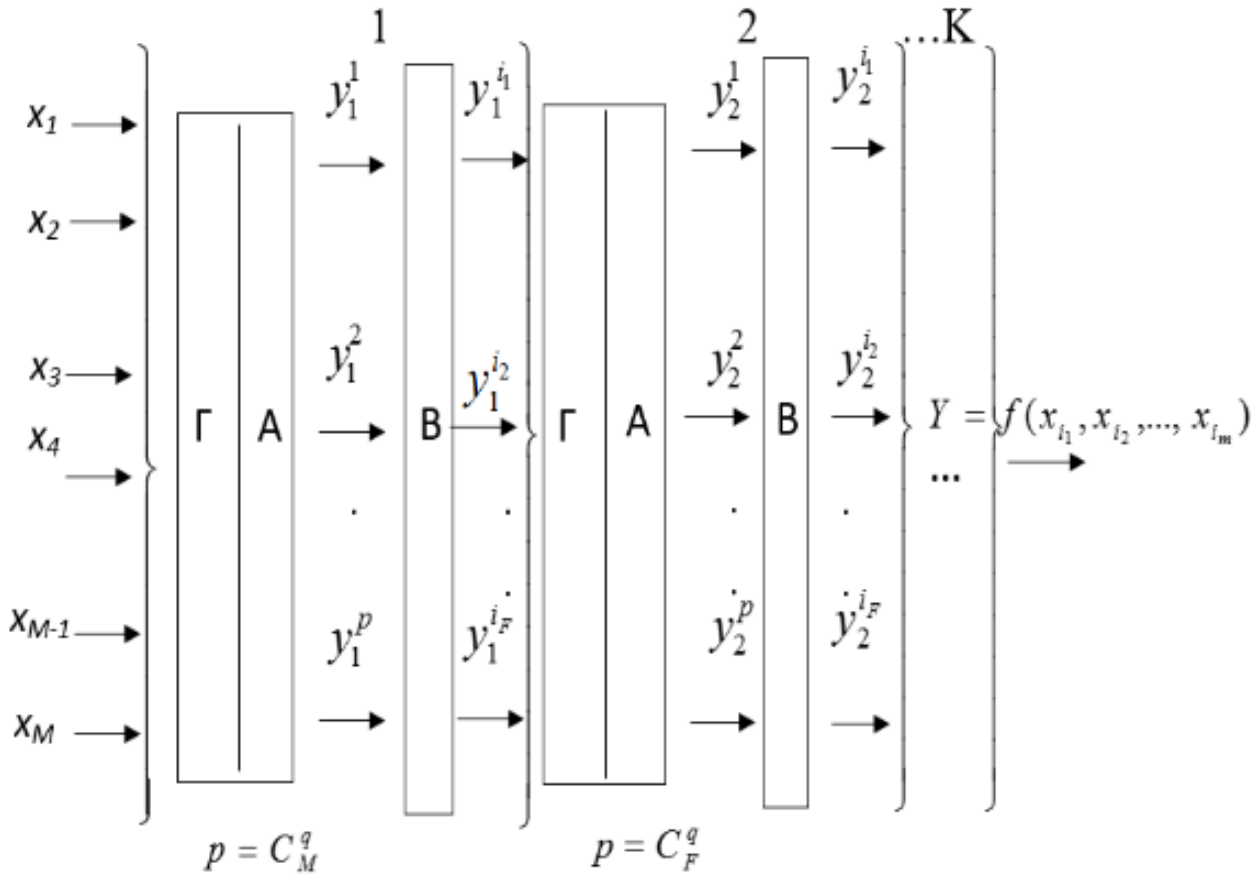


Рисунок 3.5. - Схема МГУА на прикладі багаторядного алгоритму

Тут:

Г блок - генератор часткових описів (моделей претендентів);

А блок - розрахунок коефіцієнтів часткових описів на навчальній вибірці *A*;

В блок - вибір кращих часткових описів за зовнішнім критерієм;

p - кількість часткових описів, що генеруються на кожному ряду;

F - показник свободи прийняття рішень;

K - кількість рядів селекції;

Y- модель оптимальної складності;

Коротко розглянемо роботу алгоритму по схемі на прикладі багаторядного алгоритму МГУА.

Маємо на вході алгоритму вхідний набір змінних $x_1,x_2,\ldots,x_M$. Генератор

формує певну множину моделей претендентів по деякому правилу, наприклад, поєднуючи вхідні змінні по q штук та формуючи з них квадратичну форму. Тоді на виході блоку Г маємо $p = C_M^q$ (нехай далі q=2) квадратичних форм (моделей претендентів). Для кожної з них, на множині A, шукаються найкращі оцінки вектору $\hat{a}$ параметрів. Таким чином на виході блоку А формуються p моделей претендентів вигляду

$$y_k = a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_j + a_4 x_j^2 + a_5 x_i x_j \qquad (3.23)$$

У блоці В кожна з моделей оцінюється на перевірочної вибірці B і деяка частина кращих моделей - F штук пропускається далі в наступний ідентичний ряд алгоритму. По результату роботи алгоритму на поточному ряду якість моделі оцінюють помилкою на B і на сукупній вибірці $W = A + B$ по формулі нормованої відносної середньоквадратичної помилки: Δ_B^A та Δ_W^A. Далі для наступного ряду селекції (збірка блоків ГАВ) роль вхідних змінних x грають відібрані F штук (свобода вибору) моделей y. Кількість рядів селекції нарощується до тих пір, поки зменшується помилка на B - Δ_B^A. Якщо вибірки A і B досить великі $n >> M$ і представницькі, то помилка на B може стабілізуватися, в даному випадку різниця між якістю алгоритмів розмивається і усі алгоритми дадуть приблизно однакову якість моделей (на даних без шуму). Якщо вибірки малі - $n << M$ і дані зашумлені то, як правило, матимемо виражений мінімум критерію помилки на B . Що і потрібно для подолання множинності рішень.

Відзначимо, що дана схема не дає «істинної» моделі (як її отримати при невідомих параметрах шуму?). Тут вибирається не істинна модель, і не модель, яку отримають класичні АШР на всіх точках у навчальній вибірці (на моделі Stepwise, звичайно, ми отримаємо мінімальну помилку на всій вибірці W, тому що модель налаштується під шум в рамках наданого рівня значущості). А буде отримана модель оптимальної складності. Вона буде

більш проста, ніж модель з істинною структурою і більш проста і менш точна, ніж модель отримана АШР на всіх точках W . Але, як правило, така модель має кращі помилки на вибірках С - екзамені. Такий ефект відбувається, так як ми налаштовуємо структуру моделі на свіжих точках B і, оскільки B і C належать до однієї генеральної сукупності, можливо очікувати її структурної налаштованості і на інші екзаменаційні точки C . Таким чином, програючи на B (не включаючи їх в навчальну вибірку A), ми отримуємо кращі помилки на C . Що і є «надзавданням» будь-якого завдання моделювання.

3.3.3. Основоположні принципи МГУА

Система алгоритмів, що ми розглядаємо (алгоритмічні відображення) базуються на фундаментальних принципах самоорганізації, таких як еволюція, схрещування, мутація, селекція, свобода вибору. Такі підходи тепер часто називають «біологічно інспіровані когнітивні архітектури». Розглянемо найбільш важливі з принципів МГУА [5]:

1. Принцип неостаточності рішення і свободи вибору (принцип Д.Габора)

У різних реалізаціях МГУА використовуються наступні можливості:

на кожний наступний ряд пропускається не одне, а кілька кращих рішень. Крім того, для уникнення виродження процесу використовуються випадкові схрещування з елементами попередніх рядів і обмежені мутації елементів, що вижили (були відібрані до наступного ряду);

2. Відбір по виживанню у процесі еволюції

Алгоритм реалізує (блок В) відбір структур вже готових, налаштованих на нові умови, що зустрічає модель (точки B). Ці моделі запускаються в подальшу еволюцію. І що виходить в результаті? На фіксованій множині даних рішення дається наступним твердженням:

3. Принцип зовнішнього доповнення (принцип Геделя): структура моделі вибирається при найменшому значенні зовнішнього критерію.

Відповідно до принципу зовнішнього доповнення Геделя, при

поступовому ускладненні структури моделі значення зовнішніх критеріїв спочатку зменшуються, а потім збільшуються. Таким чином, в процесі алгоритмічного відображення при поступовому збільшенні рівня складності моделі отримуємо мінімум зовнішнього критерію (зовнішнього доповнення), що і сигналізує нам про знайдену оптимальну структуру.

4.Базове твердження МГУА: Модель оптимальної складності (МОС) знаходиться на мінімумі зовнішнього критерію**.**

Тобто, найкращою структурою моделі вважається та, при якій досягається мінімум зовнішнього критерію. І її називають моделлю оптимальної складності. Те, що мінімум повинен бути, це зрозуміло, так як нулем помилка бути не може, а якась модель повинна мати мінімум.

Збіжність алгоритму можна супроводити ще такими міркуваннями.

Закономірність, яка пов'язує точки *A* і *B*, що знаходяться в просторі X, обумовлена існуванням виходу *Y*, а шум цю закономірність розмиває. До тих пір, поки у міру ускладнення моделі структура і параметри узгоджено покращують помилку, пов'язану з незашумленим виходом буде покращуватися і помилка на *B*. Це пояснюється тим, що *A* і *B* - вибірки з *W*, належать одній генеральній сукупності, з одними властивостями, зумовленими зв'язком з незашумленим виходом *Y*. Як тільки процес моделювання вичерпає можливості узгодженого налаштування на залежність між *Y* та X почнеться процес підлаштування моделі під шум: таке налаштування, в основному, йде по точкам *A*, так як саме на *A* розраховуються параметри; налаштування структурами, на *B,* більш грубий і слабкий інструмент. То оскільки точки В не беруть участі в параметричному налаштуванні на шум, то і помилка на *B* почне збільшуватися. Строгий аналітичний доказ даного механізму існує, але не будемо його тут наводити.

На рис.3.6 наведено графіки зміни помилки $\Delta_W^W(s)$ на точках робочої вибірки *W* моделі, розрахунок параметрів якої зроблений на *W*, і помилки Δ_B^A (помилка на *B* моделі, параметри якої знайдені на *A*) при меншому

(крива 1), та більшому (крива 2) рівні шуму в вихідних даних. При деякому рівні шуму, найкращою моделлю стає середнє процесу (на графіку позначено лінією МО).

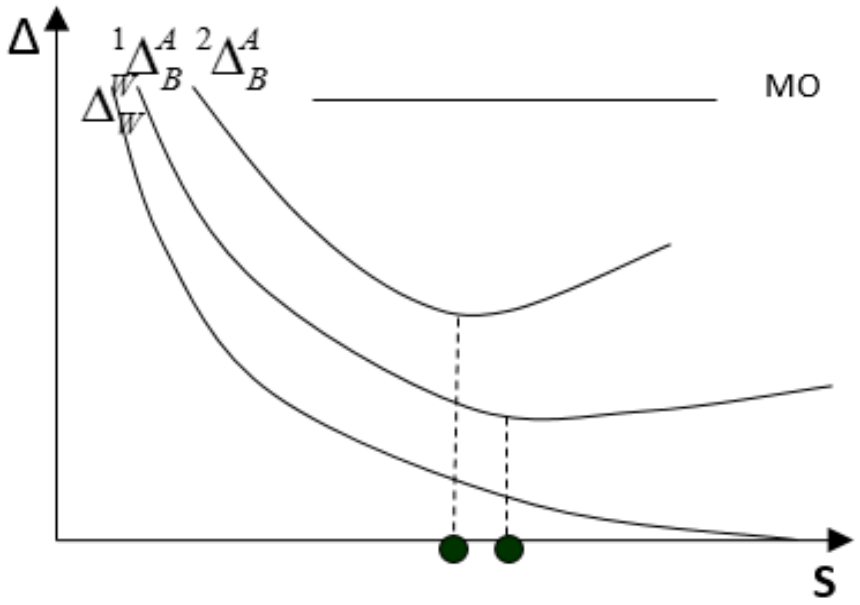


Рисунок 3.6. - Залежність критеріїв помилки моделі при різних рівнях шуму [5]

3.3.4. Властивості моделі оптимальної складності

Наведемо найбільш важливі властивості моделі оптимальної складності [5]:

1. При збільшенні рівня шуму модель оптимальної складності спрощується.

Причиною спрощення є те, що при більшому рівні шуму модель раніше почне підлаштовуватися під шум параметрами, відповідно, раніше почне погіршуватися помилка на B .

Наслідок: при певному рівні шуму середнє стає кращою моделлю.

2. Особливо важливо: Модель оптимальної структури (МОС), отримана на даних з шумом, оцінює незашумлений вихід, як правило, точніше ніж модель отримана на істинній структурі на даних з шумом з оцінками МНК по всіх точках.

"Як правило" - це значить при достатньому рівні шуму. Для кожного завдання це свій рівень, але, як правило, досить зовсім невеликого рівня шуму, щоб отримати ефективність моделі оптимальної складності для оцінки незашумленим входом перед моделлю з істинною структурою оціненої по зашумленній вибірці на всіх точках. Доказ цієї властивості досить складний, але нижче наведемо простий приклад, що пояснює, чому таке можливо.

На рис.3.7 крива 1 відповідає даним без шуму, крива 2 - на дані накладено шум у тому ж спектрі, що характеризує сигнал. Модель, як середнє (рідкий пунктир) є кращою, в сенсі суми квадратів помилки, ніж точна структура, яка за допомогою МНК своїми параметрами підлаштовується під шум (частий пунктир). Тобто, при наявності шуму модель необхідно спрощувати. І саме тому МОС корисніша в сенсі оцінки істинних, незашумлених значень виходу, ніж модель на точній структурі з МНК оцінками по всіх точках. Основний висновок з базового затвердження МГУА такий:

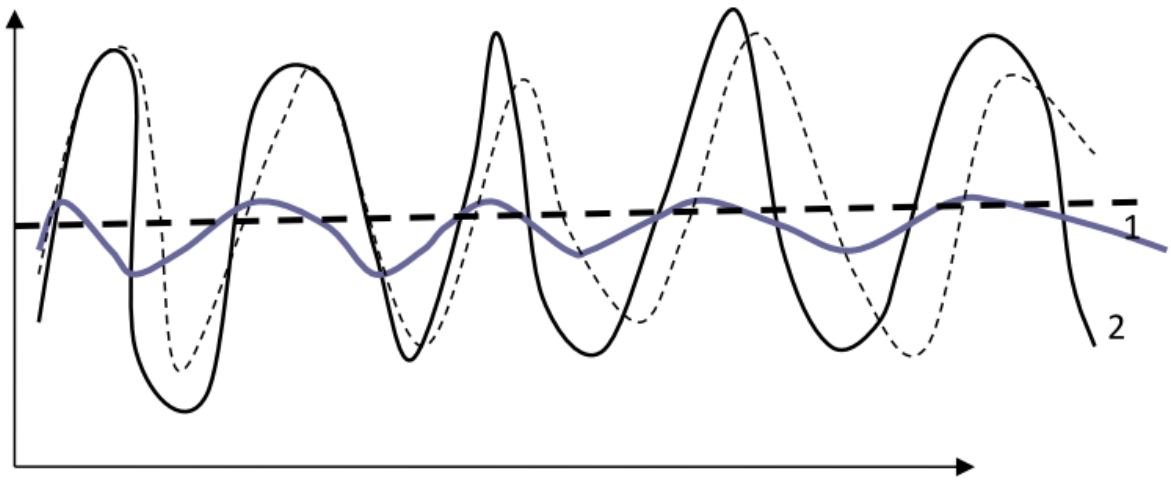


Рисунок 3.7. - Порівняння МОС і моделі істинної структури при значному рівні шуму

Вже при досить малих рівнях шумів модель оптимальної складності (маючи зміщені оцінки параметрів) оцінює незашумлений вектор виходу більш точно, ніж модель з оцінками МНК на точній структурі (які при

цьому теоретично є незміщеними)

3.4. Метод групового урахування аргументів МГУА. Алгоритми.

3.4.1. Багаторядний алгоритм МГУА

Коротко ми знайомилися з даним алгоритмом у розділі 3.4.2 при розгляді загальної схеми алгоритмів МГВА саме на прикладі багаторядного алгоритму. На цьому прикладі очевидні і явні недоліки застосування багаторядних алгоритмів для скорочення повного перебору моделей: від ряду до ряду наявні скачки складності моделі, що демонструють очевидну недосконаліть генератору структур алгоритму з точки зору селекції моделі оптимальної складності.

3.4.2. Комбінаторний алгоритм МГУА

Відзначимо, що ніякі алгоритмічні хитрощі в плані побудови різних генераторів часткових структур були б не потрібні, якби ми володіли достатніми обчислювальними ресурсами. В цьому випадку завдання пошуку оптимальної структури вирішується на заданій множині опорних функцій (скажімо поліноми) комбінаторним алгоритмом, що реалізує повний перебір підструктур із заданої повної структури поліному максимально можливого в даній задачі степеню. Але, як зазначалося вище, в даний час для реалізації повного перебору структур у реальних практичних завданнях ресурсів принципово недостатньо. Тому вся решта алгоритмів МГУА є не що інше, як різні доцільні способи скорочення повного перебору моделей, з тим, щоб намагатися не перебираючи всю множину структур, все ж відшукати серед них оптимальну або квазіоптимальну модель.

Відмітимо**,** що любий алгоритм індуктивного моделювання складається з 3-х основних частин: перша - генератор структур, друга - механізм розрахунку параметрів моделей на внутрішньому критерії, третя - відбір кращих моделей по зовнішньому критерію. Як очевидно, КОМБІ також складається з цих 3-х частин. І будь-який інший алгоритм є нічого іншого, як

багатократне (на кожному ряді або етапі) повторення цієї трійки (трійка ГАВ – див. рис. 3.5), коли вихід одного ряду/етапу є входом в наступний і відрізняються різними способами генерації (модифікації) часткових описів, внутрішніми і зовнішніми критеріями.

3.4.3. Алгоритм багаторядний, комбінаторний в вузлах

Ми знайомі у загальних рисах з 2-ма варіантами алгоритмів МГУА - багаторядний (БР) і комбінаторний (КОМБІ)**.** Кожен з них має переваги:

Для МР притаманна швидкість формування складних структур,

Для КОМБІ - знаходження глобального оптимуму якості моделі,

і недоліки :

МР формує стрибкоподібну зміну складності від ряду до ряду, а для КОМБІ притаманна вкрай мала область розмірності задачі для ефективного застосування.

Однак поєднання цих алгоритмів: багаторядний, але при цьому комбінаторний в вузлах, нівелює недоліки МР, КОМБИ та не применшує їх переваг. Алгоритм може поступово нарощувати складність моделі і при цьому в вузлах здійснюється повний перебір часткових моделей із заданого у вузлі повного поліному малої розмірності (наприклад, розмірності 2).

3.4.4. Удосконалений алгоритм Stepwise

Алгоритм реалізує багаторазову процедуру пошуку моделей претендентів згідно класичного алгоритму **Stepwise** при різних значеннях $F^{вкл}$, $F^{викл}$ (з деякої сітки множини можливих значень) та забезпечує відбір найкращої моделі за доцільно вибраним зовнішнім критерієм.

3.4.5. Релаксаційний ітераційний алгоритм МГУА з оцінками Шелудько

Нижче ми розглянемо одну з найбільш вдалих версій релаксаційних ітераційних алгоритмів МГУА (автор - Шелудько Олег Іванович), яку у період винаходу та розробки у 1975-1990 роках називали «Багаторядний спрощений алгоритм методу групового врахування аргументів»[12] . У міру розвитку МГУА розроблялася природна система класифікації алгоритмів

і з урахуванням місця даної розробки та її удосконалення в системі алгоритмів МГУА даний інструмент зараз має назву «релаксаційний ітераційний алгоритм МГУА з оцінками Шелудько» (РІА Шелудько).

Алгоритм відноситься до так названої різновиди ітераційних алгоритмів з вкладеними структурами (релаксаційний). Дійсно, для довільного ряду з номером s вирази для часткових моделей y^s виглядають як вкладені один в один описи:

$$y^s = y^{s-1} + a \cdot \hat{z} \qquad (3.24)$$

де y^{s-1} - краща модель (s -1) ряду селекції, $\hat{z}$ - деякий новий аргумент, що вводиться на s -тому ряду, a - параметр, який розраховується по МНК.

При завданні критеріїв якості $н_A$ (розрахунок параметрів) і НВСКП $_B$ (критерій оцінки структури) відповідно яким породжується модель, генератор структур забезпечує процес, що збігається на навчальній послідовності. Факт збіжності очевидний, так як в гіршому випадку, при $a = 0$, маємо $y^{s-1} = y^s$ і алгоритм дає рішення в точці стабілізації, тим самим забезпечуючи монотонність процесу. Внутрішню збіжность даного конкретного алгоритму показано на завершення його опису.

Нижче позначення *y, z, x* вживаються двояко, як змінні моделювання в рівняннях моделі, і як відповідні вектори в просторі об'єктів (точок) на малюнках і у формулах розрахунку параметрів. У тексті позначення супроводжуються відповідним терміном змінна (аргумент) або вектор.

Вхідні дані алгоритму є значення вхідних змінних $x_j \; j = \overline{1,m}$ і вихідної змінної y , що представлені на рис.3.8 матрицею даних *X* і вектором виходу y , m - кількість вхідних змінних.

y	x_1	x_2	…	x_m
y_1	x_{11}	x_{12}	…	x_{1m}
y_2	x_{21}	x_{22}	…	x_{2m}
…	…	…	…	…
y_n	x_{n1}	x_{n2}	…	x_{nm}

Рисунок 3.8. - Вхідні дані алгоритму

Розглядається вибірка з n точок - множина W, що поділена на навчальну A та перевірочну (тестову) B послідовності даних.

1. Підготовчий ряд алгоритму: формується розширення множини вхідних змінних - змінні виду $1/x$, $\sqrt[3]{x}$, $1/\sqrt[3]{x}$ для кожної x_j $j=\overline{1,m}$. Тоді розширеною множиною T вхідних змінних для алгоритму, буде сумарна множина змінних: X: $\{x_j\}\forall j$, $j=\overline{1,m}$, та їх функцій Ω: $\left\{1/x_j, \sqrt[3]{x_j}, 1/\sqrt[3]{x_j}\right\}, \forall j,\ \ j=\overline{1,m}$

Позначимо $\mathrm{T}=\Omega\cup X$. Формуємо множину узагальнених вхідних змінних алгоритму $Z=\{z_k\}$, як множину добутків нульового, першого другого, і ін. порядків від змінних початкової вхідної множини T**:**

$$Z=\left\{z_k\right\}=\mathrm{T}\cup\mathrm{T}\otimes\mathrm{T}\cup\ldots\cup\mathrm{T}\otimes\mathrm{T}\ldots\otimes\mathrm{T}=\left\{t_i\right\}\cup\left\{t_{i_1}\cdot t_{i_2}\right\}\cup\ldots\cup\left\{t_{i_1}\cdot t_{i_2}\cdot\ldots\cdot t_{i_{kr}}\right\} \qquad (3.25)$$

При формуванні часткових моделей

$$y^{s+1}=y^s+a_k\cdot\hat{z}_k \qquad (3.26)$$

послідовно вводять у модель ортогоналізований варіант $\hat{z}_k$ узагальнених змінних z_k, взятих з множини $Z = \{z_k\}$, де кількість множників $\{t_{i_1} \cdot t_{i_2} \cdot \ldots \cdot t_{i_{kr}}\}$ у доданку моделі z_k, називають максимальним числом корекцій аргументу - число kr, параметр алгоритму. Кількість елементів (потужність) даної множини $\{z_k\}$ позначимо через *K*.

На довільному s-тому ряду алгоритму формується *F* найбільш перспективних шляхів отримання найкращих описів і по завершенню селекції s-того ряду залишається один найкращий опис. Тому в класичному розумінні теорії МГУА в даній версії алгоритму ступінь свободи *F*=1. У більш пізніх версіях даного алгоритму передбачається можливість формування *F* кращих рішень для передачі на наступний ряд.

Стандартні версії релаксаційних алгоритмів при розрахунку параметрів моделі y^s з двома аргументами (y^{s-1} та $\hat{z}_k$) передбачають розрахунок трьох параметрів - a_{k0}, a_{k1}, a_{k2} моделі

$$y^s = a_{k0} + a_{k1} \cdot y^{s-1} + a_{k2} \cdot z_k \qquad (3.27)$$

Однак, якщо є можливість без суттєвої втрати точності моделі зменшити кількість параметрів, що розраховуються, це приведе до істотного збільшення швидкодії алгоритму.

Зменшимо розмірність вектора параметрів, що шукаємо. Нагадаємо, що розмірность вектора параметрів - це одночасно розмірність системи нормальних рівнянь, або що те ж , розмірність інформаційної матриці $X^T X$. У нашому випадку, кожний частковий опис формується (з урахуванням вільного члена) із змінних $x_1 = 1$, $x_2 = y^{s-1}$ та $x_3 = \hat{z}_k$. Зменшення розмірності приведе, в тому числі, і до зменшення помилки розрахунку параметрів за відомою нам формулою $\hat{a} = (X^T X)^{-1} X^T Y$. Нижче з цією метою

приведемо (3.27) до виду (3.26). Відповідний підхід до розрахунку параметрів розгорнутої форми рівняння (3.26) отримав назву розрахунку оцінок Шелудько.

2. Етап центрування вхідних змінних

Для представлення (3.27) в еквівалентну двопараметричну форму проведемо центрування вхідних (для кожної моделі) узагальнених змінних z_k і вихідну змінну y. Змінні центруються тільки за даними навчальної вибірки *A,* так як тут мова йде про механізм розрахунку параметрів:

$$\tilde{z}_{ij} = (z_{ij} - z_j^{\grave{A}}) \quad \tilde{y}_i = (y_i - y^{\grave{A}}), \quad j = \overline{1,K}, \quad i = \overline{1,n} \tag{3.28}$$

де $\bar{z}_j^A, \bar{y}^A$ - значення середніх на вибірці А:

$$\bar{y}^A = \frac{1}{N_A}\sum_{i=1}^{N_A} y_i \ , \ \ \bar{z}_j^A = \frac{1}{N_A}\sum_{i=1}^{N_A} z_{ji} \tag{3.29}$$

Тоді за участю центрованих змінних рівняння моделей будуть фігурувати вже в двопараметричному вигляді

$$\tilde{y}^s = a_{k1} \cdot \tilde{y}^{s-1} + a_{k2} \cdot \tilde{z}_k \tag{3.30}$$

для деякої k-тої моделі на s-тому ряду алгоритму. Після знаходження оптимальної структури можемо повернутися до форми з трьома параметрами. Маємо

$$(y^s - \bar{y}) = a_{k1} \cdot (y^{s-1} - \bar{y}^{s-1}) + a_{k2} \cdot (z_k - \bar{z}_k) \tag{3.31}$$

і з огляду на $y = a_{k0} + a_{k1} y^{s-1} + a_{k2} z_{k2}$ можемо знайти для (3.33) параметр

$$a_{k0} = \bar{y} - a_{k1}\bar{y}^{s-1} - a_{k2}\bar{z}_k \qquad (3.32)$$

Надалі перепозначення змінних опустимо, маючи на увазі, що нижче застосовуються центровані змінні.

3. Обгрунування підходу до розрахунку параметрів моделі.

Описана нижче процедура формування моделі першого і наступних рядів дозволяє знаходити модель у вигляді

$$y^s = y^{s-1} + a_k \hat{z}_k \qquad (3.33)$$

де y^{s-1} - краща модель (s -1) ряду селекції , $\hat{z}_k$ - ортогоналізована до y^{s-1} складова аргументу z_k, a_k - вже єдиний параметр моделі, який шукаємо методом найменших квадратів. На даному етапі ми не стверджуємо про еквівалентність форм моделей (3.27) та (3.33), проте форма (3.33) з одного боку, дозволяє знизити розмірність розрахунку завдання до єдиного параметра, з іншого, отримати *ортогоналізований* опис. Даний підхід до розрахунку параметрів моделі при розгортанні опису до вхідних узагальнених змінних z отримав назву розрахунок оцінок Шелудько.

Нагадаємо, що параметри, розраховані на ортогональних (незалежних) вхідних аргументах, найточніше розраховуються в умовах шумів. Для цього варіанту розрахунку площина аргументів, куди опускають проекцію у просторі узагальнених змінних розташована найбільш «стійко» щодо шуму в даних. Дійсно, кут між векторами $\mathbf{y}^{s-1}$ та $\hat{\mathbf{z}}_k$, якими ця площина нами визначена - 90^0. є найбільш стійким відносно впливу шуму на її положення Таким чином шум даних впливає на положення площини в найменшому ступені і порахована модель має найменшу похибка пов'язану з наявністю шуму. Отже застосований механізм ортогоналізації дозволяє отримувати найкращі, з точки зору завадостійкості, оцінки параметрів.

4.Формування першого ряду моделей алгоритму

На першому ряду розглядаються всі одночленні моделі наближення об'єкта

$$y^{1} = a \cdot z^{1} \tag{3.34}$$

де z^{1} перебираються з множини $z^{1} \in Z$. Коефіцієнти a часткових моделей першого ряду знаходимо по МНК на точках навчання (множині точок *A),* вимагаючи наближення моделями (3.40) виходу y .

Виведення формули МНК для одновимірного випадку проводиться за тим же принципом, що виводиться загальна формула МНК з умовної системи рівнянь: необхідно наблизити модельний вектор в просторі точок $\mathbf{y}$ його моделлю $\mathbf{y}^{1}$:

$$\mathbf{y} \overset{\Delta}{=} \mathbf{y}^{1} \tag{3.35}$$

Тому з (3.34) запишемо:

$$\mathbf{y} \overset{\Delta}{=} a\mathbf{z}^{1} \tag{3.36}$$

Звідси, домноживши зліва кожну сторону (3.36) на вектор рядок $(\mathbf{z}^{1})^{T}$ (далі знак транспонування будемо опускати, застосувавши знак скалярного добудку векторів), отримаємо для кожної k-тої моделі претендента

$$\mathbf{z}_{k}^{1} \cdot \mathbf{y} = a_{k}^{1} \mathbf{z}_{k}^{1} \cdot \mathbf{z}_{k}^{1} \tag{3.37}$$

або

$$\hat{a}_k^1 = \frac{\mathbf{z}_k^1 \cdot \mathbf{y}}{\mathbf{z}_k^1 \cdot \mathbf{z}_k^1}, k = \overline{1, K} \tag{3.38}$$

де « • » - знак скалярного добутку векторів.

Формулу (3.38) можемо записати у вигляді

$$\hat{a}_k^1 = \frac{\sum_{i \in I_A} z_{k_i}^1 y_i}{\sum_{i \in I_A} (z_{k_i}^1)^2}, \quad k = \overline{1, K} \tag{3.39}$$

де I_A - множина індексів навчальних точок A вибірки даних.

Параметр $\hat{a}_k^1$, що знайдено - коефіцієнт проекції виходу **y** на напрям вектора обраної узагальненої змінної z_k^1 (рис. 3.9). Визначивши за значенням критерію селекції (про нього мова піде далі) кращу модель першого ряду, продовжимо генерацію часткових описів на наступних рядах.

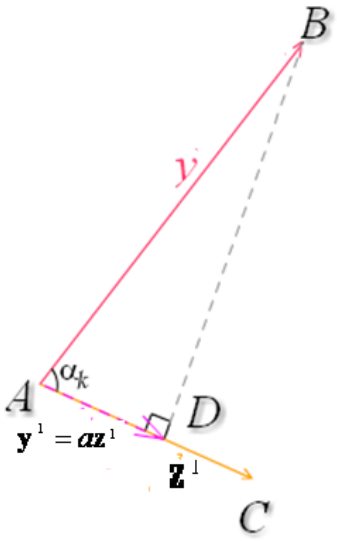


Рисунок 3.9. - Визначення параметру моделі a (параметру проекції виходу **y**) на напрямі узагальненого аргументу $\mathbf{z}^1$ на 1-му ряду алгоритму

5. Формування довільного s-того ряду алгоритму.

Далі шукаємо ортогоналізовані часткові моделі у вигляді

$$\mathbf{y}_k^s = \mathbf{y}^{s-1} + a_k^s \hat{\mathbf{z}}_k^s, \quad s > 1 \tag{3.40}$$

тут s - індекс ряду, $\mathbf{z}_k^s$ - сформований, як кращий з F прорахованих шляхів знаходження узагальненої змінної шляхом утворення добутку з розширеної множини вхідних змінних T**,** при цьому $\hat{\mathbf{z}}_k^s$ - це ортогональна вектору $\mathbf{y}^{s-1}$ складова вектора $\mathbf{z}_k^s$, $\mathbf{y}^{s-1}$ - найкраща модель попереднього ряду.

За умови центрування змінних за середнім значенням в часткових моделях (35) вільний член дорівнює нулю, а коефіцієнт при третьому члені моделі

$$a_k = \frac{\sum\limits_{i \in i_A} z_{k_i} \varphi_i^{s-1}}{\sum\limits_{i \in I_A} (z_{k_i})^2} \tag{3.41}$$

де вектор $\varphi^{s-1} = \mathbf{y} - \mathbf{y}^{s-1}$ є невязка моделі попереднього ряду, а $\hat{\mathbf{z}}_k^s$ - це ортогональна $\mathbf{y}^{s-1}$ складова вектору $\mathbf{z}_k^s$.

Ортогональна добавка $\hat{\mathbf{z}}_k^s$ до $\mathbf{y}^{s-1}$ з рівняння (3.40) зв’язана з $\mathbf{z}_k^s$ векторною залежністю

$$\hat{\mathbf{z}}_k^s = \mathbf{z}_k^s - A_k \mathbf{y}^{s-1} \tag{3.42}$$

Дісно, зверніть увагу на прямокутній трикутник на рис. 3.10, де коефіцієнт A_k визначається з умови ортогональності

$$\mathbf{y}^{s-1} \cdot \hat{\mathbf{z}} = 0 \quad \text{чи} \quad \sum_{i \in I_A} y_i^{s-1} z_{k_i} = 0, \tag{3.43}$$

звідси маємо

$$A_k = \frac{\sum_{i \in i_A} y_i^{s-1} z_{k_i}}{\sum_{i \in i_A} (y_i^{s-1})^2} \qquad (3.44)$$

Зміст проведеного перетворення у тому, що вихідний вектор $\mathbf{z}_k$ розкладається за двома напрямками – за напрямку вектора $\mathbf{y}^{s-1}$, і у ортогональному йому напрямку - $\hat{\mathbf{z}}_k^s$: $\mathbf{z}_k = \hat{\mathbf{z}}_k + A_k \mathbf{y}^{s-1}$ (рис. 3.10). Тому в подальших розрахунках (визначенні a_k) бере участь тільки його частина $\hat{\mathbf{z}}_k^s$, що ортогональна вектору кращої моделі попереднього ряду $\mathbf{y}^{s-1}$.

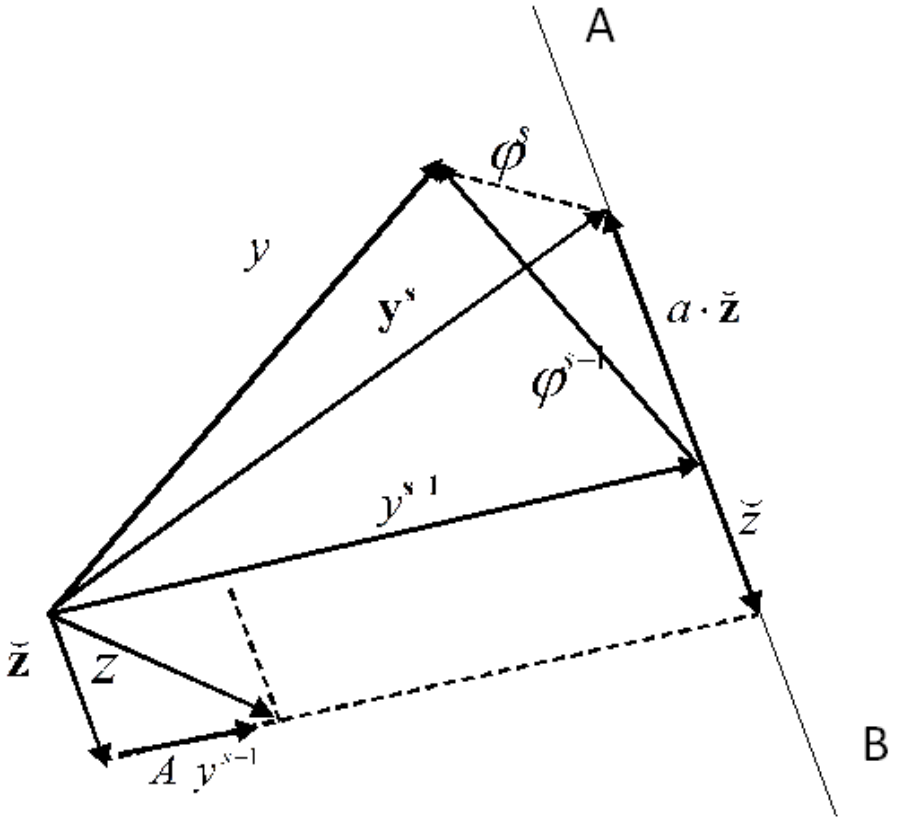


Рисунок 3.10. - Визначення параметру проекції виходу **y** на напрямі узагальненого аргументу $\mathbf{z}^s$ на s-му ряду алгоритму

Геометрично, вектор $\mathbf{y}^s$ одержується шляхом проекції вектору **y**

(вектор нев'язки φ^s - позначено пунктиром) на напрямок АВ, що є продовженням вектора $\hat{\mathbf{z}}_k$. Я очевидно, процедура проекції є процедурою розрахунку коефіцієту проекції a_k за МНК у площину аргументів ($\mathbf{y}^{s-1}$, $\mathbf{z}_k$) на напрям у ній, ортоганальний $\mathbf{y}^{s-1}$.

6. Внутрішня збіжність алгоритму за моделями y^s до вектору виходу y .

Розглянемо факт внутрішньої збіжності алгоритму - збіжності на навчальній послідовності даних, що є необхідною умовою застосування будь-якого алгоритму структурного синтезу.

Збіжність послідовності y^1, y^2 ,.... , y^{s-1}, y^s ,..... до вихідного вектору y стає очевидною з розгляду прямокутного трикутника, що формується векторами φ^{s-1}, φ^s та вектором $a \cdot \breve{z}$ (рис 3.10). Тут, як вказувалося вище, φ^{s-1}, φ^s - нев'язки кращих моделей $\varphi^{s-1} = y - y^{s-1}$, $\varphi^s = y - y^s$ відповідного ряду. Якщо врахувати, що в прямокутному трикутнику φ^{s-1}, φ^s , $a \cdot \breve{z}$, вектор φ^{s-1} - гипотенуза, а вектор φ^s - катет, то стає очевидним незростаючий характер послідовності невязок, $\varphi^1, \varphi^2, ... \varphi^{s-1}, \varphi^s, ...$ а отже і збіжність послідовності векторів $y^1\ y^2$ .... $y^{s-1}\ y^s$ ..... на навчальної послідовності до вихідного вектору y.

7. Критерій селекції алгоритму РІА Шелудько

Вибір кращих моделей на кожному ряду алгоритму здійснюється відповідно до значень критерію селекції, який є складовим критерієм, що складається з комбінованого критерію точності і критерію незміщенності рішень.

$$K_k = \beta(\alpha \cdot \Delta_A^A + (1-\alpha)\Delta_B^A) + (1-\beta)\frac{\left|\Delta_A^A - \Delta_B^A\right|}{\left|\Delta_A^A + \Delta_B^A\right|} \qquad (3.45)$$

тут

$$\Delta = \frac{\sum_{i=1}^{N_w}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_w}(y_i - \bar{y})^2},$$

і Δ_A^A - помилка на вибірці *A,* що обчислена на моделі, параметри якої оцінювалися на вибірці *A,* Δ_B^A - помилка моделі на вибірці *B,* що порахована на моделі, параметри якої оцінювалися на вибірці *A,* α - вага критерію точності на навчальній вибірці, $(1-\alpha)$- вага критерію точності на перевірочної вибірці, β - вага комбінованого критерію точності в складеному критерії, $(1-\beta)$ - вага критерію незміщенності рішень в складеному критерії, $\alpha, \beta \le 1$.

Робота алгоритму завершується після досягнення мінімуму критерію селекції. Відповідна модель є моделлю оптимальної складності. Алгоритм включає можливість здійснювати 2 типу розбиття я вибірки даних на навчальну і перевірочну послідовно: перші N_A - навчання, наступні N_B - перевірка (для задач прогнозування), або випадковим чином (для задач аппроксимації).

Зауважимо, що у версії РІА БАКСУ критерій має вигляд

$$K_k = (\beta-1)(\alpha \cdot \Delta_A^A + (1-\alpha)\Delta_B^A) + \beta\frac{\left|\Delta_A^A - \Delta_B^A\right|}{\left|\Delta_A^A + \Delta_B^A\right|} \qquad (3.46)$$

Підведемо підсумок розгляду алгоритму. Позитивні сторони релаксаційного ітераційного алгоритму на основі оцінок Шелудько:

1.Сниження розмірності розрахунку параметрів моделі при переборі структур до одиниці на порядки покращує швидкодію алгоритму в порівнянні з будь-якими минулими і сучасними алгоритмами структурного

синтезу.

2. Ортогоналізація узагальнених аргументів (перетворення z_k в $\hat{z}_k$) відносно найкращого рішення попереднього ряду y^{s-1} дозволяє отримувати ортогоналізований опис з найкращими, з точки зору завадостійкості, оцінками параметрів.

Плата за зазначені істотні переваги - не оптимальність одержуваних оцінок параметрів щодо оцінок МНК для тієї ж розгорнутої структури. Оцінки Шелудько збігаються до них тільки в міру збільшення рядів алгоритму. Однак в більшості практичних завдань переваги алгоритму з надлишком перекривають зазначений недолік.

3.5. Класифікація методів індуктивного моделювання

3.5.1. Схема та місце алгоритмів МГУА у множині методів індуктивного моделювання

Досить багато розробок, що отримано науковими дослідниками світової спільноти відповідають принципам МГУА (згадані вище підходи з явним і неявним штрафом за складність моделі), тому вони та МГУА можуть бути об'єднані деякою загальною назвою. В Україні використовують термін *методи індуктивного моделювання.*

Нижче наведена схема функціональних блоків індуктивного моделювання (генератори, зовнішні і внутрішні критерії) яка відображає різноманіття інструментів вирішення різних класів задач моделювання. У схемі на рис.3.11 показано і місце результатів, близьких за ідеологією до МГУА.

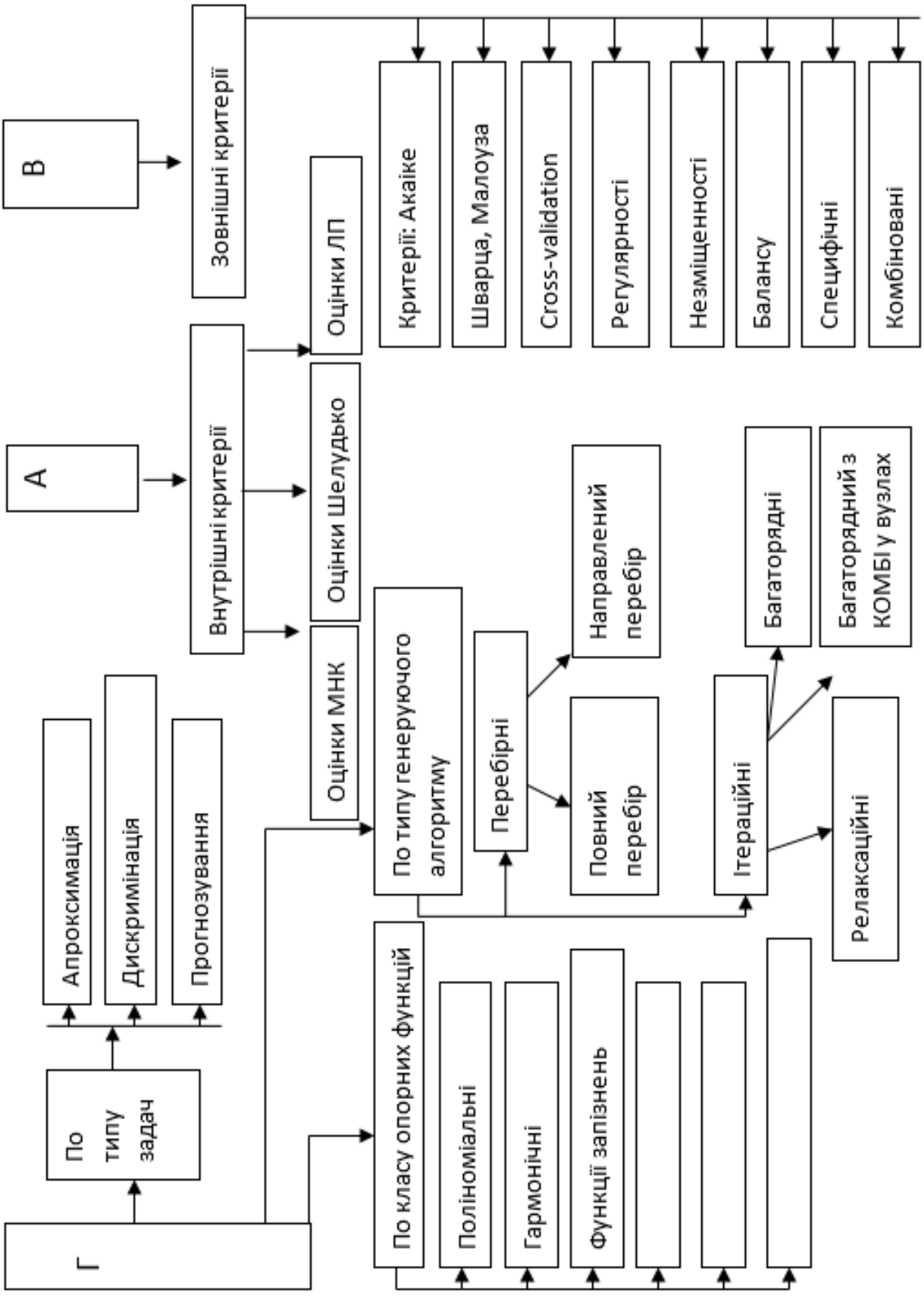


Рисунок 3.11. - Класифікація алгоритмів і методів індуктивного моделювання

Класифікацію природно проводити відповідно до блоків загальної схеми роботи таких алгоритмів - блоків Г, А і В :

1. Варіації генераторів структур моделей, що застосовуються дозволяють класифікувати алгоритми через

1.1 . Тип вирішуваних завдань:

1.1.1. Апроксимації,

1.1.2. Дискримінації (діагностики, класифікації),

1.1.3. Прогнозування.

1.2 . Клас застосованих опорних функцій:

1.2.1. Клас поліноміальних опорних функцій,

1.2.2. Клас гармонійних опорних функцій,

1.2.3. Клас опорних функцій з запізнюваннями.

1.3. Тип генеруючого алгоритму:

1.3.1. Перебірні алгоритми (з фіксованим порядком перебору):

1.3.1.1. Повний перебір (комбінаторний алгоритм),

1.3.1.2. Спрямований перебір: МАЛТІ - багатоетапний селекційно-комбінаторний алгоритм, або як варіант, рекурентна версі,я

1.3.2. Ітераційні алгоритми:

1.3.2.1. Релаксаційні ітераційні алгоритми: РІА Шелудько, БАКСУ, РІАПДС (РІА на повному дереві структур),

1.3.2.2. Багаторядні (з квадратичними описами),

1.3.3. Змішаного типу:

1.3.3.1. Багаторядний, комбінаторний у вузлах,

1.3.3.2. Удосконалений Stepwise.

2. Розрахунок параметрів часткових описів проводиться відповідно до внутрішнього критерію алгоритму, найбільш відомі:

2.1. Оцінки МНК (з підкласом рекурентного оцінювання по МНК),

2.2. Оцінки Шелудько,

2.3. Оцінки задач матпрограммування (симлекс метод, тощо).

3. Класифікація алгоритмів за зовнішнім критерієм селекції структур:

3.1 Критерій Малоуза,

3.2 Критерій Акаіке,

3.3. Критерій Шварца,

3. . Критерій "cross-validation",

3.5. Критерій регулярності,

3.6. Критерії незміщенності (рішень та/або параметрів),

3.7. Критерії балансу,

3.8 . Специфічні критерії: критерій гладкості, кризовий критерій,

3.9 . Комбіновані критерії.

3.5.2. Опис структурної схеми методів індуктивного моделювання

Нижче додано опис тільки для тих елементів схеми, що не згадувалися раніше по тексту.

1.Генератори. Відмінності в генераторах структур часткових моделей забезпечують можливість вирішувати різні класи задач.

1.1. Генератори для класів задач:

1.1.1. Апроксимації,

1.1.2. Дискримінації (діагностики, класифікації),

1.1.3. Прогнозування.

Відмітимо, що на сьогодні особливої спеціалізації немає - окремі генератори можуть застосовуватися в кожній задачі – наприклад, механізм послідовного ускладнення моделі степеневими аргументами і перехресний механізм формування часткових моделей може застосовуватися і для 1.1.1, 1.1.2, 1.1.3. Особливості класів задач можуть бути враховані застосуванням різних класів опорних функцій. Але в принципі така спеціалізація можлива.

1.2. Генератори для класів опорних функцій:

1.2.1. Клас поліноміальних опорних функцій

1.2.2. Клас гармонійних опорних функцій

1.2.3. Клас опорних функцій з запізнюваннями

Для апроксимації і в задачах класифікації цілком підійдуть

поліноміальні опорні функції, а в задачах прогнозу частіше потрібні гармонійні та функції з запізнюваннями.

На клас алгоритму моделювання найбільший вплив має тип генератору

1.3. Тип генеруючого алгоритму:

1.3.1. Перебірні алгоритми (з фіксованим порядком перебору),

1.3.1.1. Повний перебір (комбінаторний алгоритм КОМБІ),

1.3.1.2 Спрямований перебір (МАЛТІ – багатоетапний комбінаторно- селекційний алгоритм). Тут реалізуються послідовні ускладнення моделі $y^s = a_0 + a_1 z_1 + \ldots + a_{s-1} z_{s-1} + a_s \cdot z_s$, коли кожен раз перебираються найкращі F її продовжень і перераховуються її параметри.

1.3.2. Ітераційні алгоритми

1.3.2.1. Релаксаційні ітераційні алгоритми РІА: РІА Шелудько, БАКСУ, РІАПДС (РІА на повному дереві структур) [13]. Властивість релаксаційності визначає вкладеність структур на шляху ускладнення моделей, різні алгоритми відрізняються різними механізмами формування структурної частини, що додається на кожному ряду багаторядної процедури та механізмом розрахунку оцінок параметрів. Наприклад, РІА Шелудько реалізує послідовне ускладнення структури на двочленній моделі $y^s = y^{s-1} + a \cdot \hat{z}_s$, де розрахунок оцінки параметру a враховує ортогоналізацію узагальненої змінної z_s до y^{s-1}.

1.3.2.2. Багаторядні (з квадратичними описами)

1.3.3 . Змішаного типу: багаторядний, з КОМБІ у вузлах

2. Внутрішні критерії: розрахунок параметрів часткових описів проводиться відповідно до внутрішнього критерію алгоритму, найбільш

відомі:

2.1 Оцінки МНК. Стандартний запис для оцінок МНК $\mathbf{a} = (X^T X)^{-1} X^T Y$, де X - матриця об'єкт-властивості, Y - вектор виходу.

2.2 Рекурентне оцінювання за МНК

При розрахунку параметрів моделей

$$y^s = a_0 + a_1 z_1 + ... + a_{s-1} z_{s-1} + a_s \cdot z_s \qquad (3.47)$$

кожне нове додавання аргументу приводить до повного перерахунку заново всього набору коефіцієнтів за формулою $\mathbf{a} = (Z^T Z)^{-1} Z^T Y$, причому відомо, що час розрахунку зворотної матриці $A^{-1} = (Z^T Z)^{-1}$ пропорційний m^3, де m - порядок інформаційної матриці $Z^T Z$. Однак, можливо на порядок, до m^2 знизити витрати ресурсу, якщо для розрахунку на n-тому кроці матриці A_n^{-1} використовувати знання розрахованої раніше, на (n -1) кроці, матриці A_{n-1}^{-1} . Реалізує цю можливість метод "обрамлення"[14]. Якщо позначимо в структурі матриці A_n елементи як

$$A_n = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1,n-1} & a_{1,n} \\ a_{21} & a_{22} & ... & a_{2,n-1} & a_{2,n} \\ ... & ... & ... & ... & ... \\ a_{n-1,1} & a_{n-1,2} & ... & a_{n-1,n-1} & a_{n-1,n} \\ a_{n,1} & a_{n,2} & ... & a_{n,n-1} & a_{n,n} \end{bmatrix} = \begin{bmatrix} A_{n-1} & u_n \\ v_n & a_{n,n} \end{bmatrix}, \qquad (3.48)$$

то зворотну їй матрицю A_n^{-1} шукаємо як:

$$A_n^{-1} = \begin{bmatrix} A_{n-1}^{-1} + \dfrac{A_{n-1}^{-1} u_n v_n A_{n-1}^{-1}}{a_n} & -\dfrac{A_{n-1}^{-1} u_n}{a_n} \\ -\dfrac{v_n A_{n-1}^{-1}}{a_n} & \dfrac{1}{a_n} \end{bmatrix} \tag{3.49}$$

$$a_n = a_{nn} - v_n A_{n-1}^{-1} u_n \tag{3.50}$$

Детальніше з технологією, що використовуються у МГУА для прискорення розрахунку можна познайомиться у книзі [5] стор. 78-69.

2.3. Оцінки Шелудько [12]: розраховуються для моделей у вигляді

$$y^s = y^{s-1} + a \cdot \hat{z}_s \tag{3.51}$$

Параметр a тут розраховується за МНК, але для ортогоналізованних аргументів. Параметри такої розгорнутої щодо x моделі

$$y^s = a_0 + a_1 z_1 + ... + a_{s-1} z_{s-1} + a_s \cdot z_s \tag{3.52}$$

сходяться до оцінок МНК, у міру збільшення кількості рядів алгоритму. Даний недолік в практичних завданнях з лишком компенсується багатократною перевагою у швидкості розрахунку на відповідному рівні складності моделі. Тому оцінки Шелудько за принципом швидкість/точність конкурентоздатні з прямим розрахунком, а в задачах великої розмірності у нього конкурентів немає.

3. Зовнішні критерії селекції структур.

Нижче наведена коротка характеристика лише для тих критеріїв, що не розглядалися вище

3.4. Критерій "cross-validation"

Існує багато реалізацій принципу перехрестної перевірки, у тому числі,

коли ролі навчальної та перевірочної послідовності міняються місцями.

3.6. Критерії незміщенності [5].

Критерій орієнтовано на незміщеність рішень та/або параметрів моделі

3.7. Критерії балансу [5].

Критерій орієнтований на збереження властивостей, що виконувались у минулому та повинні виконуватись у майбутньому

3.8.Специфічні критерії:

Приклади: критерій гладкості, що у дискретному варіанті реалізується, як сума модулей перших різниць та кризовий критерій, що визначаний як кількість збігів перемін знаків похідної процесу та моделі

3.9. Комбіновані критерії

Критерії реалізують будь-яку доцільну комбінацію окремих зовнішніх критеріїв, чи реалізують послідовний порядок їх застосування.

3.6. Контрольні питання

1. Наведіть варіанти постановки задачі структурно-параметричного синтезу для лінійної по параметрам регресії

2. Обгрунтуйте основні проблеми структурно-параметричного синтезу

3. Розтлумачте, яким чином обгрунтувати застосування методу моделювання модельним експериментом

4. Надайте характеристику методам індуктивного моделювання з явним штрафом за складність моделі.

5. Надайте характеристику методам індуктивного моделювання з неявним штрафом за складність моделі.

6. Надайте рзгорнуту характеристику МГУА

7. Розтлумачте поняття зовнішнього критерія

8. Поясність основоположні принципи МГУА

9. Обгрунтуйте властивості моделі оптимальної складності

10. Поясніть роботу алгоритмів МГУА на прикладі комбінаторного та

багаторядного алгоритму.

11. Розтлумачте переваги багаторядного алгоритму, комбінаторного у вузлах та удосконаленого Stepwise

12. Поясніть роботу релаксаційного ітераційного алгоритму МГУА з оцінками Шелудько (РІА Шелудько), та обгрунтуйте його переваги.

15. Розгляньте схему класифікації методів індуктивного моделювання та поясніть спільні та відмінні риси окремих підходів.

РОЗДІЛ 4. ВИРІШЕННЯ НЕТРАДИЦІЙНИХ ЗАВДАНЬ МОДЕЛЮВАННЯ

4.1. Класифікація об’єктів, заданих множинами багатовимірних спостережень

4.1.1. Введення у завдання класифікації множин.

Природним розширенням завдання розпізнавання образів, де об'єкт заданий одним багатовимірним спостереженням, - позначимо надалі таку постановку завдання як (*), є постановка задачі, де об'єкт задано множиною багатовимірних спостережень, - позначимо надалі таку постановку як (**). Практичних застовувань для даної задачі безліч – від завдань класифікації зображень (спеціальний випадок – розпізнавання обличь) до діагностичних медичних завдань, де вхідними даними є результати моніторингу спостережень за пацієнтом. Окремим корисним застосуванням цього завдання є спосіб радикального поліпшення результатів класифікації об'єктів шляхом переходу від характеристики об'єкта одним багатовимірним спостереженням до характеристики об'єкта множиною спостережень, проведених при різних умовах.

Справді, якщо " портрет " об'єкта, як одне багатовимірне спостереження недостатньо інформативний, виникає добре відома проблема неоднозначності розв'язання задачі класифікації. Труднощі розпізнавання стають ще очевиднішими, якщо значення ознак об'єкта можуть змінюватися залежно від величини деякого неконтрольованого параметра чи, у загальному випадку, від стану середовища. Проблема виникає через можливість часткового перетину областей значень ознак у вихідному просторі вимірюваних змінних для об'єктів різних класів при різних станах середовища, що призводить до очевидної неоднозначності результату класифікації.

Можливим виходом із ситуації є опис об'єкта не однією, а певною

кількістю вимірювань, проведених для різних станів середовища. Такий набір вимірювань дозволяє більш точно описати об'єкт класифікації, як певного набору його станів у вихідному багатовимірному просторі. Як наслідок, ми звертаємось до нового класу завдань класифікації (**), де об'єкти визначаються наборами багатовимірних спостережень. Прикладом такого підходу є система оцінки функціонального стану серцево-судинної системи (ССС) людини за вимірами артеріального тиску (АСТ -АДТ) та пульсу. Одноразовий вимір може лише констатувати поточне відхилення роботи ССС, проте дослідження зміни даних параметрів при різних навантаженнях дає достатню інформацію для набагато детальніших діагностичних висновків.

Проте існуючий, добре розроблений апарат розпізнавання образів орієнтований виключно постановку завдання класифікації багатовимірних спостережень і не може бути застосований до розв'язання завдання класифікації множин. Ситуацію погіршує і те, що діагностичні системи, як правило, є основною ланкою інформаційних систем підтримки прийняття медичних рішень. Умови функціонування таких медичних систем характеризуються високим ступенем неоднозначності поведінки суб'єктів задачі та наявністю прихованих змінних, знання яких необхідне для прийняття управлінських рішень. Саме для таких завдань прийняття рішень насамперед необхідно розробляти підходи для вирішення розширеної постановки задачі класифікації об'єктів (**).

4.1.2. Існуючі підходи до розв'язання задачі класифікації множин.

Підходи до розв'язання задачі за умовою (**) можна поділити на дві групи: безпосереднє розв'язання задачі (**) та підходи, що розглядають зведення завдання у формулюванні (**) до завдання у постановці (*).

Перші підходи до розв'язання задач класу (**) були розроблені для двох типів об'єктів: об'єктів, заданих одномірною ознакою $z(t)$ (рис. 4.1) та об'єктів, заданих значеннями багатовимірної ознаки **z** на площині зображення (рис. 4.2).

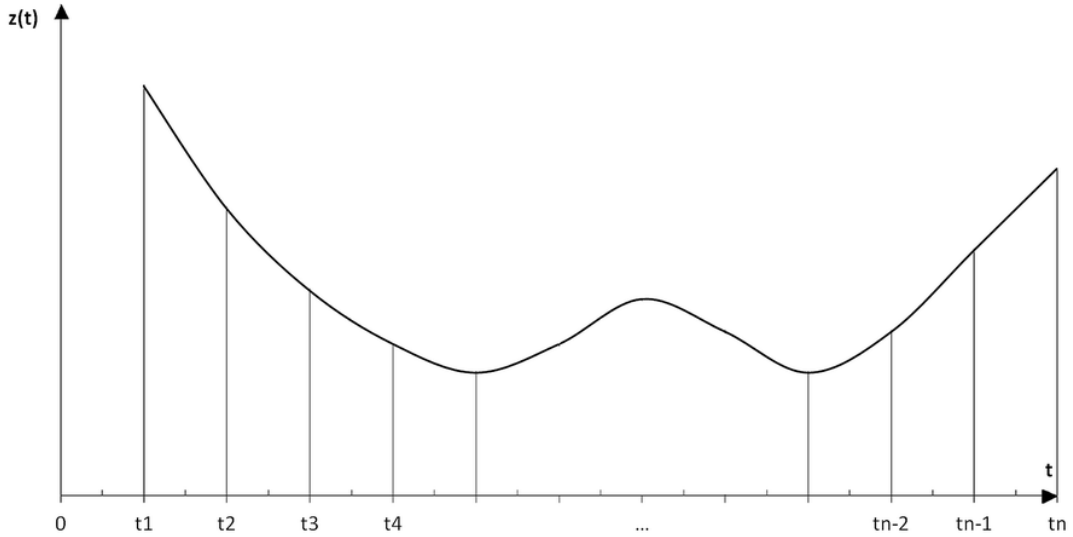


Рисунок 4.1. - До завдання класифікації кривих

Щоб перезадати завдання класифікації у формулюванні (**) для першого типу об'єкту, де криві визначаються значеннями одномірної ознаки *z* у відліках часу $t_1,t,\ldots,t_n$, у формулювання (*), введемо нові ознаки: ознака z_1 буде визначатися значеннями $z(t_1)$, ознака z_2, значеннями $z(t_2)$, ..., ознака z_n, значеннями $z(t_n)$.Таким чином, ми перевизначили об'єкт-криву, що була задана множиною точок $z(t_1), z(t_2),\ldots, z(t_n)$ у одновимірному просторі *z*, до багатовимірного спостереження вектора **z**=($z_1, z_2,\ldots, z_n$). Тепер до класифікації кривих можно застосувати методи класифікації багатовимірних спостережень.

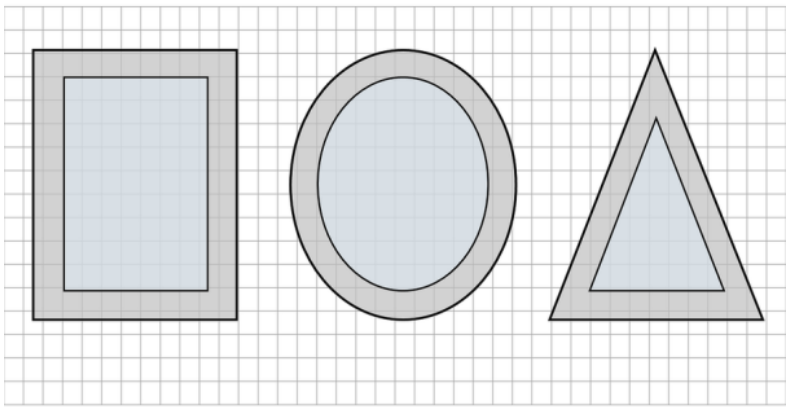


Рисунок 4.2. - До завдання класифікаціі плоских форм

Для другого типу об’єктів – плоских форм (рис.4.2), введемо припущення, що кожна з них розміщується у прямокутниках фіксованого розміру (і відповідно з фіксованою кількістю пікселів). Тоді фігури (об’єкти класифікації) у формулюванні (*) визначаються множиною значень вектора **z** $=(z_1,z_2,\ldots,z_m)$ у кожному пікселі, що належить об’єкту. Щоб перезадати завдання класифікації у формулюванні (**), введемо змінні z_{ij}, що визначаються, як різні по *i* ознаки зображення z, $i=1,\ldots,m$ у точці об'єкта з координатами заданого пікселя $j=1,\ldots,n$. Ознаки інтерпретуються, як випадкові величини, завдання пошуку класифікатора зводиться до отримання розподілів значень ознак за класами та обчислення дискримінантної функції, що дозволяє розділити ці об'єкти [15].

Позитивний досвід вирішення такого завдання шляхом порівняння функцій розподілу для різних зображень у низькорозмірному випадку наведений у роботах [16] та [17]. Однак необхідність урахування більшої кількості залежних ознак вимагає створення багатовимірної щільності з розмірністю m*n, де m - розмірність вектора ознак, а n - дискретність представлення об'єкту. Складності такого підходу добре відомі, крім того, необхідний обсяг навчальної інформації, необхідний створення достатньо точних для якісної класифікації багатовимірних щільностей, як правило, відсутня. Такі обставини знижують ефективність інструментів функцій щільності у вихідному просторі ознак класифікації об'єктів. С іншими частковими підходами рішення завдання класифікації множин можна, познайомиться у роботах [18,19].

Розглянемо докладніше обставини виникнення завдання (**) у медицині. Набір, що характеризує об'єкт, може бути результатом регістрації параметрів стану пацієнта при багатовимірному моніторингу або у процесі активного експерименту. Діагностика стану ССС людини шляхом аналізу множинних вимірювань тиску та пульсу, отриманих у різних умовах (навантажувальна крива) є ілюстрацією другого випадку. Іншим прикладом завдання у

постановці (**) у медичних застосуваннях є діагностичний аналіз МРТ, КР, УЗ зображень органів людини. Об'єкти в цьому разі визначаються відповідним набором точок (пікселів). Кожна точка має характеристики, що містяться у форматі даних DICOM. У загальному випадку це означає, що об'єкти, що підлягають класифікації, задаються матрицею ознак, в якій об'єкт задається не одним рядком, як звичайно у випадку (*), а деякою кількістю. При цьому кількість таких рядків може бути різною для кожного об'єкту. Кожен такий рядок є одним багатовимірним спостереженням. Така сама матриця формується, коли клінічні параметри пацієнта реєструються у різних умовах чи є даними моніторингу. Умови вимірювання можуть бути невідомі, що означає можливу наявність прихованих змінних, що впливають об'єкт. Приклад двох таких об'єктів, позначених наборами « 1 » і « 2 » у просторі ознак x_1, x_2 показаний рис. 4.3.

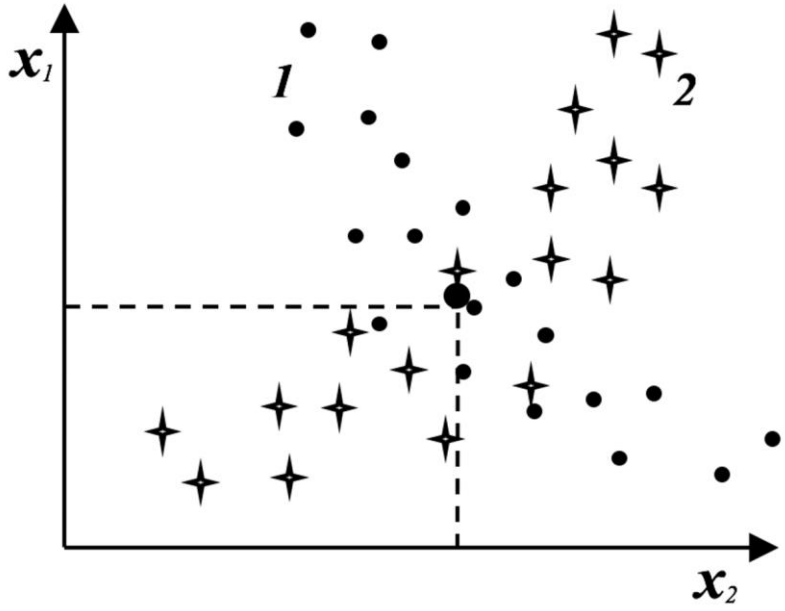


Рисунок 4.3. - Приклад розташування спостережень різних класів об'єктів у просторі вихідних ознак

Як бачимо розділити об’єкти-множини неможливо. Особливістю даного

класу завдань є те, що множини вимірювань, що представляють об'єкти, можуть перетинатися (як на рис. 3) з точністю до тотожності деяких рядкових підмножин матриці ознак, що належать об'єктам різних класів. Ось кілька актуальних робіт, в яких розглядається даний клас завдань: автори [20], [21] розглядають діагностику та прогноз захворювання шляхом повторних спостережень або моніторингу параметрів пацієнта , автори [22], [23] розглядають задачу спостереження та аутентифікації людини, автори [24], [25] розглядають задачу класифікації патологій з медичних зображень.

Розглянутий клас задач природно називати задачами класифікації множин (Set Classification Problem - SCP). Згідно з Юнгом і Цяо [26]: «Завдання класифікації множин виникають, коли класифікація заснована на аналізі наборів спостережень, а не на окремих спостереженнях» (стор. 1). «Необхідне правило класифікації витягується з наборів спостережень, а також має бути спроможним передбачити єдину мітку класу для кожного нового набору спостережень» (стор. 3).

Серед робіт, що безпосередньо вирішують SCP, можна вказати [22], [23], [27], які вирішують завдання аутентифікації. Цей підклас завдань відрізняється тим, що в окремий клас представлена конкретна людина або набір образів, що відноситься до цієї людини. Маючи низку інших наборів зображень, необхідно встановити ідентифікацію обличчя людини, що ідентифікується, або визначити приналежність об’єкту до інших існуючих класів. Методи засновані на тому чи іншому перетворенні наборів зображень до деякої більш компактної форми, що порівнюється з формою іншого об’єкту класу.

У роботі [22], де об’єктом є фіксована послідовність представлення людини з поворотом зліва направо, кожне зображення перетворюється на ряд характеристик (eigenfaces) і у цьому просторі формується векторна траєкторія. Порівняння траєкторій дає підстави для віднесення об'єкта до класу чи для відмови від класифікації. У роботах [23],[27] застосовані щільності ймовірності розподілу ознак, - це ще одна компактна форма

представлення образів. Підходи у ряді випадків показали свою працездатність завдяки умовам формування зображень [22] і методам перетворення простору ознак [23], [27]. Вказані обставини дозволили звести завдання до малорозмірного випадку.

У роботах [28], [29] розробляють ефективний підхід до згортання порівнюваної інформації. Тут вирішується проблема міри близькості коваріаційних матриць образів, що лежать на симетричному позитивно визначеному рімановому різноманітті. Це дозволяє отримувати хороші результати у завданнях, де можливе формування ознак з яскраво вираженою лінійною залежністю, як основи для відмінності класів.

Підводячи підсумок аналізу існуючих методів вирішення завдань класифікації множин, можна дійти висновку, що існуючі підходи вирішують тільки часткові, окремі випадки цього завдання й у разі довільної нелінійної залежності між ознаками об'єктів (що ми можемо спостерігати у наборах спостережень) необхідно запропонувати узагальнений підхід , що дозволяє вирішити поставлене завдання на основі аналізу структур і параметрів цих залежностей. Інші роботи, близькі до пропонованого підходу, обговорюються у наступних розділах.

4.1.3. Формальна постановка задачі

Маємо факт існування деякого набору класів $D = \{\overline{D}^1, \overline{D}^2, ..., \overline{D}^K\}$. Ці класи є кінцевими або нескінченними наборами об'єктів : $\overline{D}^1 = \{\overline{d}^1\}$, $\overline{D}^2 = \{\overline{d}^2\}$,…, $\overline{D}^K = \{\overline{d}^K\}$. Відомо, що набори $\overline{D}^i \cap \overline{D}^j = \varnothing$ коли $i \neq j$.

Класи $\overline{D}^1, \overline{D}^2, ..., \overline{D}^K$ задаються, як їх апроксимації $D^1 \subset \overline{D}^1$, $D^2 \subset \overline{D}^2$ … , $D^K \subset \overline{D}^K$ через усічені множини об'єктів, що належать цим класам:

$$D^1 = \{d^1\} \subset \{\overline{d}^1\},\ D^2 = \{d^2\} \subset \{\overline{d}^2\},\ ..,\ D^K = \{d^K\} \subset \{\overline{d}^K\},$$

де вказані множини:

$$\{d^1\} = \{d_1^1, ..., d_{n_1}^1\},\ \{d^2\} = \{d_1^2, ..., d_{n_2}^2\},\ \ldots,\ \{d^K\} = \{d_1^K, ..., d_{n_K}^K\}.$$

Очевидно, що якщо вірно $\overline{D}^i \cap \overline{D}^j = \varnothing$ то вірно і $D^i \cap D^j = \varnothing$, при $i \neq j$.

Передбачається, що об'єкти класифікації d задаються відповідними множинами векторів вимірювань $\{\mathbf{z}\}_d$ у кінцевовимірному просторі $Z \subset R^M$ змінних $(z_1,...,z_M)$, де вектори $\mathbf{z}$ утворюють множини (області) $\overline{Z}^1, \overline{Z}^2, ..., \overline{Z}^K$, що відповідають класам $\overline{D}^1, \overline{D}^2, ..., \overline{D}^K$. Як зазначалося вище, у просторі Z можливий частковий перетин множин $\{\mathbf{z}\}_d$, що належать різним класам (як на рис. 4. 3), що призводить до можливого часткового перетину множин (областей) $\overline{Z}^i$. Необхідно розробити найкраще правило класифікації для поділу об'єктів вихідних множин $\{\overline{D}^1, \overline{D}^2, ..., \overline{D}^K\}$з урахуванням навчальних підмножин об'єктів d із різних класів $Z^1 \subset \overline{Z}^1$, $Z^2 \subset \overline{Z}^2$, …, $Z^K \subset \overline{Z}^K$.

4.1.4. Узагальнений підхід до вирішення задачі класифікації множин

Розглянемо інші можливі шляхи вирішення проблеми, що мають більш загальний характер. Спочатку формалізуємо дані на вході нашого завдання. Нехай у просторі $Z \subset R^M$ задано навчальні підмножини $Z^1 \subset \overline{Z}^1$, $Z^2 \subset \overline{Z}^2$, . .. , $Z^K \subset \overline{Z}^K$ об'єктів $\{d_j^i\}$, де i - індекс класу, j - індекс об'єкта класу. Кожен об'єкт d_j^i визначається у просторі Z підмножинами $Z_{d_j^i}$ багатовимірних спостережень (векторів-рядків $\mathbf{z}_t$) матриці ознак M, $Z_{d_j^i} = \{\mathbf{z}_t\}_{d_j^i}$. Нехай дані у матриці впорядковані за класами, позначимо n_k - кількість об'єктів у класі k, $n = \sum_{k=1}^{K} n_k$ - кількість об'єктів (груп строк) матриці. Відмітимо далі кількість спостережень у кожному класі. Якщо це число k в кожному об'єкті класу однаково і дорівнює N_{W^k}, то $N_k = n_k \cdot N_{W^k}$. Якщо кількість спостережень в об'єктах d_i^k різна і рівна $N_{W_i^k}$, то $N_k = \sum_{i=1}^{n_k} N_{W_i^k}$. Тут індекси W^k та W_i^k позначають навчальні підмножини спостережень у відповідних класах, та окремих об'єктах класу k, відповідно.

Загальна кількість спостережень (рядків) у матриці даних дорівнює $N = \sum_{\mathrm{k}=1}^{\mathrm{K}} N_K = \sum_{k=1}^{K} \sum_{i=1}^{n_k} N_{W_i^k}$. Істотним далі є те, що для кожного об'єкта d_i^k відомо не одне, а деяке підмножина спостережень $Z_{d_i^{\mathrm{k}}} = \left\{ \mathbf{z}_j \right\}_{d_i^k}$, тут $j = \sum_{p=1}^{k-1} N_p + \sum_{q=1}^{i-1} N_{W_q^k} + 1, \ldots, \sum_{p=1}^{k-1} N_p + \sum_{q=1}^{i} N_q^k$, де i - номер об'єкта в класі k .

Істотно також те, що в загальному випадку області існування об'єктів d_i і d_j, що представлені в навчальних підмножинах векторами $Z_{d_i} = \left\{ \mathbf{z}_t \right\}_{t \in d_i}$, $Z_{d_j} = \left\{ \mathbf{z}_t \right\}_{t \in d_j}$, $Z_{d_i}, Z_{d_j} \subset Z$ можуть частково перетинатися, і при цьому ці об'єкти можуть належати різним класам. Тоді можна розглянути наступні два способи перетворення постановки (**) до задачі класифікації в умовах (*). Перший з них полягає в тому, щоб знайти для об'єктів d_{j}^k згортку $z_{d_{\mathrm{j}}^k}^{\theta} = \sum_{\mathbf{z} \in \mathrm{W}(k,j)} \theta \cdot \mathbf{z}$ підмножини спостережень $Z_{d_j^k} = \{\mathbf{z}\}_{d_j^k}$ таким чином, щоб вона тотожно визначала об'єкт d_j у своєму класі k у вихідному просторі ознак $z_1, \ldots, z_M$. За таким визначенням згортки має слідувати умова визначення параметрів θ для досягання найкращої класифікації об'єкта в даному класі згортки. Найпростіший варіант згортки - представлення об'єктів центроїдами їх множин. Два набори (точки та зірки), що представляють два об'єкти різних класів, показані на рис. 4.3. Приклад показує низьку ефективність представлення об'єктів центроїдами: центри збігаються за ознаками, і об'єкти у такому представлення не можуть бути розрізненими. Щодо можливості побудови класифікаторів або поділу таких об'єктів складними границями на основі вихідних ознак: врахуємо, що окремі спостереження об’єктів, що можуть як найближче збігатися у деяких об’єктів різних класів у будь-якому іншому просторі після перетворення вихідних ознак також збігатимуться. Отже, немає класифікаторів чи границь будь-якого рівня складності, які б диференціювати такі об'єкти у просторі вихідних ознак.

Другий підхід враховує, що залежність вихідних ознак і наявність

набору вимірювань для кожного об'єкта дає можливість моделювання характерних зв'язків, що лежать в основі явищ, що вивчаються. Особливості структури та значення параметрів цих моделей можуть бути використані як основа системи класифікації. Таким чином, сенс підходу полягає у вирішенні задачі класифікації у просторі параметрів моделей об'єктів класифікації.

Сама ідея вирішувати різноманітні завдання у просторі параметрів моделі об'єктів не нова. Цей підхід застосований до завдань ідентифікації параметрів регресійних моделей, які такі, що найкраще вирішують завдання апроксимації чи прогнозування. Завдання стиснення інформації аналогічні до цієї ідеї. Те саме стосується і класичних підходів до вирішення задачі класифікації в постановці (*). Продемонструємо можливість застосування того самого принципу в задачі КМ .

Нехай об'єкт d , заданий набором точок $Z_d = \{\mathbf{z}\}_d$, що відображають невідому характеристику об'єкта $f_d(z_1,...,z_M) = 0$ у вихідному просторі ознак $z_1,...,z_M$. Передбачається, що таких характеристик має бути стільки ж, скільки й об'єктів:

$$f_d(z_1,z_2,...,z_M) = 0, \quad d = 1,...,n \ . \tag{4.1}$$

Розглянемо випадок, коли серед вихідних змінних $z_1,...,z_M$ можна вибрати вихідну змінну y. Нехай X означає новий простір ознак узагальнених змінних $x_1,x_2,...,x_m$, отриманий з вихідного простору $Z \subset R^M$. Передбачається можливість представлення характеристик (1), заснованих на множинах векторів $Z_d = \{\mathbf{z}\}_d$ у вигляді лінійних згорток по узагальненим змінним $x_1,x_2,...,x_m$. Тоді характеристики (4.1) можна знайти у такому вигляді:

$$y = a_0 + \sum_{i=1}^{m} a_i \varphi_i(z) = a_0 + \sum_{i=1}^{m} a_i x_i \ , \tag{4.2}$$

де m означає розмірність простору узагальнених змінних x_i, $i = 1,..., m$ в якому представлені об'єкти d. Тепер розв'язання задачі класифікації можна перенести у простір параметрів $\mathrm{A} \subset R^m$, так як значення векторних параметрів $\mathbf{a} = (a_0, a_1,..., a_m)$ характеристик $y = a_0 + \sum_{i=1}^{m} a_i x_i$ будуть визначати об'єкти в цьому новому просторі. Окремі вектори $\mathbf{a}_d$ однозначно визначають відповідні об'єкти d через відсутність підмножин $Z_{d_j} = \{\mathbf{z}\}_{d_j}$, що повністю збігаються. Отже маємо можливість класифікувати об'єкти d, як вектори $\mathbf{a}_d$ у просторі параметрів А.

Заперечення щодо використання такого підходу можуть бути пов'язані з можливим порушенням гіпотези компактності та супутніми проблемами, пов'язаними з вибором міри близькості у просторі А при вирішенні задачі класифікації. Однак, як буде показано, цю обставину можна врахувати, вибравши відповідний зовнішній критерій для моделювання (4.2). Приклад, поданий на рис.4.3, рис.4. 4, рис.4.5 показує можливу ефективність такого підходу.

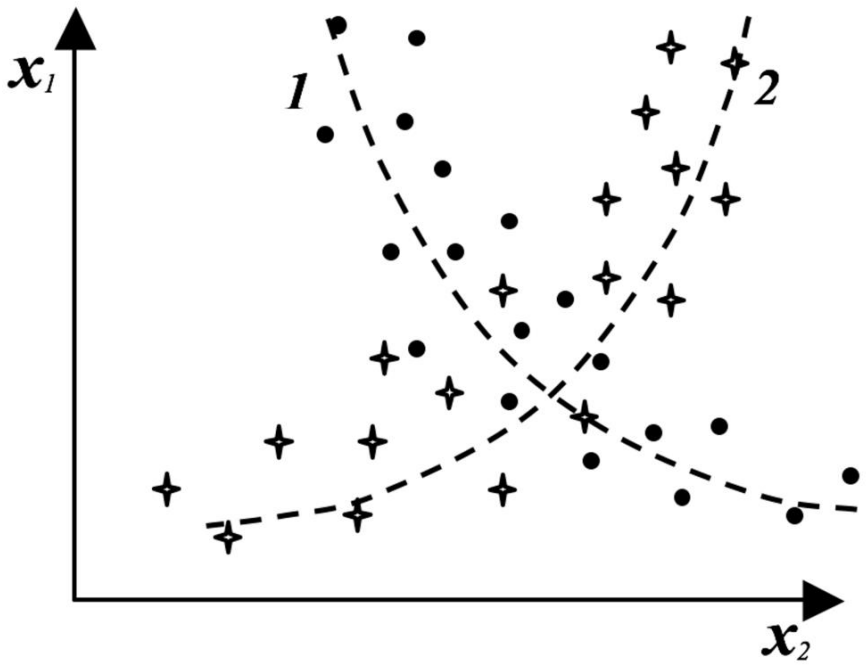


Рисунок 4.4. - Моделі об’єктів у просторі початкових ознак

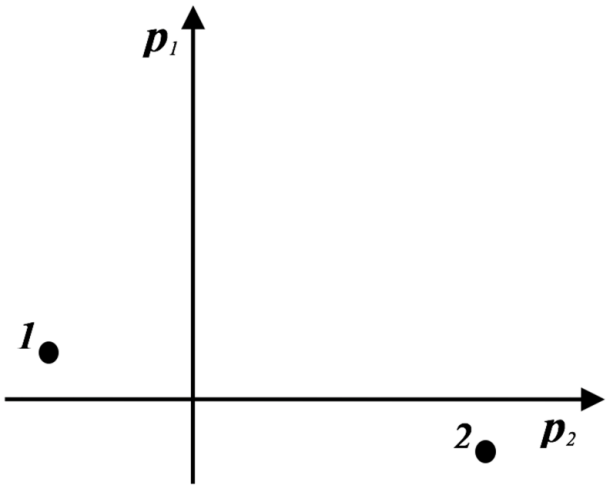


Рисунок 4.5. - Представлення об’єктів у просторі параметрів моделі

Очевидно, об'єкти 1,2, що визначаються своїми наборами вимірювань, нероздільні у просторі вихідних ознак (рис. 4.3). Але після створення моделей об’єктів по відповідним наборам даних (рис. 4.4), у просторі параметрів моделей (параметрів гіперболи $x_1 = 1/(p_1 x_1 + p_2) + k$) зазначені об'єкти вже ефективно розділені (рис. 4.5). Ефективність розділення тим вища, що більше відрізняється структура та/або значення параметрів побудованих моделей. Отже, при вирішенні задачі класифікації у постановці (**) може бути доцільним перейти із простору вихідних ознак у простір параметрів моделей об'єктів.

Вкажемо нижче роботи, які найближче підійшли до реалізації цієї ідеї. У роботі [30] показано вирішення задачі діагностики інфаркту серцевого м'яза на підставі ознак, які були отримані, як параметри розкладання сигналу електрокардіограми до ортогонального дискретного ряду Кравчука. У роботі [20] авторам вдалося здійснити перехід до параметричного простору за допомогою досить простих математичних засобів. Тут розглянуто проблему прогнозування гострого гіпотензивного епізоду під час спостереження кардіологічного хворого. Контекст проблеми докладно описаний у [31]. Спостереження проводилося за хворими, які перебувають у відділенні реанімації щодо основного захворювання. У деяких пацієнтів за період спостереження був хоч один гострий гіпотензивний епізод. Інші пацієнти

проблем із тиском не мали. Ці групи пацієнтів сформували навчальні вибірки 2-х класів з високим та низьким ризиком розвитку загрозливої гіпертензії. Дискретність вимірювання тиску даних часового ряду, що використовується, становила 15 хвилин. За отриманими наборами вимірювань будували лінійні регресії систолічного, діастолічного тиску та пульсу. Потім у чотиривимірному просторі параметрів отриманих моделей методом опорних векторів будувалася система класифікації. Отримано досить високий показник правильної класифікації - 93%. Слід зазначити, що завдання вирішувалося за досить грубого поділу на класи (два класи), що дозволило отримати якісну класифікацію на основі значень параметрів лінійної регресії. Однак більш детальний підхід до оцінки функціонального стану серцево-судинної системи (5 і більше класів) вимагає побудови нелінійних моделей залежності параметрів пульсу, систолічного та діастолічного тиску [21]. Тому бажано запропонувати більш загальний підхід, що дозволяє шукати нелінійні структури, та орієнтувати моделювання об'єктів на якість розв'язання задачі класифікації множини. Такі вимоги можуть бути задоволені у межах теорії алгоритмів методу групового урахування аргументів [5], [13] . Далі наведемо засади 2-х варіантів реалізації підходу до вирішення завдань КМ [19]:

- Випадок, коли існує достатньо точна єдина спільна структура моделей для всіх об'єктів класифікації (всіх класів). Тоді моделі об’єктів створюються в рамках такої єдиної структури, при цьому параметри моделі відображають характерні риси класів. Виконання задачі моделювання структурного синтезу спрямоване на пошук саме такої структури для всіх об'єктів. В результаті кожен об'єкт може бути представлений точкою в просторі параметрів отриманих моделей. Саме в цьому просторі вирішуватиметься завдання класифікації, причому тепер може бути застосований будь-який з існуючих методів розв'язання задачі у постановці (*);
- Структурні відмінності класів суттєві, отже неможливо відображати їх особливі якості у межах єдиної спільної структури моделей об'єктів

класифікації. Тоді ми змушені створювати моделі у кожному класі окремо на основі оптимальної структури для кожного класу. При цьому оптимальність можемо сформулювати у термінах:

а) сукупної точності моделей об'єктів класу та

б) якості відокремлення об'єктів класу від усіх інших класів.

Завдання класифікації може вирішуватися відповідно до точності представлення об'єкта моделлю для кожної знайденої структури класів, або може вирішуватись класифікаторами у просторі параметрів структур класів за принципом «один проти всіх», або класифікаторам у просторі параметрів «об'єднаної» структури.

4.1.5. Алгоритми пошуку найкращої спільної структури моделей об’єктів

Наведемо 2 алгоритми пошуку найкращої спільної структури для перетворення простору вирішення задачі класифікації множин:

- Комбінаторний алгоритм методу групового урахування аргументів для перетворення простору спостережень (GMDH combinatorial algorithm for transforming the observational space - COMBITOS GMDH) орієнтований на отримання найменшої сукупної помилки всіх моделей. Алгоритм гарантує глобальний оптимум під час пошуку структури моделей, але необхідність повного перебору всіх структур-кандидатів обмежує раціональне використання алгоритму кількістю аргументів трохи більше 6-9. При вищих вимірах слід використовувати багаторядні чи багатоетапні алгоритми [5].
- Ітераційний алгоритм перетворення простору спостережень за МГУА (OSCIA GMDH - Observation Space Conversion GMDH Iteration Algorithm), що реалізує скорочений перебір структур.

4.1.5.1. Комбінаторний алгоритм для перетворення простору спостережень за МГУА (COMBITOS GMDH)

Завдання алгоритму полягає в тому, щоб знайти єдину спільну для всіх об’єктів класифікації структуру (нелінійний базис $\varphi_i(\mathbf{z}) = x_i,\ i = 1,..., m$), що

найкраще представляє у певному сенсі залежності

$$y_d = a_{d_0} + \sum_{i=1}^{m} a_{d_i} \varphi_i(z) = a_{d_0} + \sum_{i=1}^{m} a_{d_i} x_i \,,\ d = 1,..., n \qquad (4.3)$$

з експериментальних даних $z_d = \{\mathbf{z}\}_d$. Теорія методу групового урахування аргументів формалізує поняття «найкращих сенсів» у зовнішніх критеріях та обгрунтовує їх використання при структурно-параметричному синтезі моделей оптимальної складності. Під оптимальністю тут розуміється досягнення найкращого значення зовнішнього критерію в процесі перебору структур при синтезі моделі [5].

Відповідно до [32] адаптація алгоритму COMBI для вирішення завдання класифікації множин вимагає:

- визначити зовнішній критерій алгоритму знаходження найкращої єдиної спільної структури для моделей об'єктів усіх класів $d = 1,...,n$;
- вибрати раціональний підхід до розрахунку параметрів конкретних моделей, щоб мінімізувати час розрахунку оптимальної конструкції;
- визначити порядок перебору структур моделей.

Вибір зовнішнього критерію

Параметри моделі розраховуються на навчальній вибірці даних, найкраща структура вибирається на основі мінімуму зовнішнього критерію відповідно до принципів МГУА [5] . Це означає, що така єдина спільна структура $\varphi_i(z) = x_i, i = 1,..., m$ шукається для всіх $d = 1,...,n$ моделей, параметри яких $a_{d_0},.., a_{d_m}$, $d = 1,...,n$ розраховуються за навчальною вибіркою, і вони в сенсі зовнішнього критерію будуть найбільш точно представляти всі моделі об'єктів (4.3).

До кожного об'єкту $d = d(k,i)$, де k – номер класу, i – номер об'єкту в класі k, вказується набір його робочих спостережень $W(k,i) = A(k,i) + B(k,i)$, де

$A(k,i)$, $B(k,i)$ – навчальна і перевірочна вибірки спостережень об'єкта i у класі k. Відповідно, $W(k) = A(k) + B(k)$, $A(k)$, $B(k)$ - робоча, навчальна та перевірочна вибірки спостережень об'єктів класу k і $W = A + B$, A, B - набори спостережень у робочій, навчальній та перевірочній вибірках об'єктів усіх класів, відповідно.

Далі будемо використовувати одну із найпростіших форм зовнішнього критерію, оскільки слабким місцем алгоритмів COMBI є загальний обсяг обчислень. Ця форма є скоригований критерій точності:

$$Cr = \alpha \cdot \Delta_A^{\gamma} + (1-\alpha) \cdot \Delta_B^{\gamma}, \qquad (4.4)$$

де α – вага помилок навчальної вибірки, $(1-\alpha)$ – вага помилок перевірочної вибірки.

Критерій помилки, розрахований для всіх моделей на навчальній вибірці спостережень :

$$\Delta_A^{\gamma} = \left(\frac{\sum_k^K \gamma_k \sum_j^{n_k} \sum_{i \in A(k,j)} (Y_{kji} - \sum_q^M x_{iq} a_{kiq}^A)^2}{\sum_{i \in A} (Y_i)^2}\right)^{-1/2}, \qquad (4.5)$$

тут Δ_A^{γ} зважується у класах з коефіцієнтами γ_k.

Критерій помилки моделей, параметри яких обраховані на навчальних вибірках, та якість яких оцінена на спостереженнях перевірочної вибірки:

$$\Delta_B^{\gamma} = \left(\frac{\sum_k^K \gamma_k \sum_j^{n_k} \sum_{i \in B(k,j)} (Y_{kji} - \sum_q^M x_{iq} a_{kiq}^A)^2}{\sum_{i \in B} (Y_i)^2}\right)^{-1/2}. \qquad (4.6)$$

Тут Δ_B^{γ} зважується в класах з коефіцієнтами γ_k, k - індекс класу, j - індекс об'єкта в класі, i пробігає всі спостереження об'єкту j класу k на вибірках A для (4.5) або вибірках B для (4.6) відповідно. У розрахунках застосовуємо центровані значення змінних, тому немає потреби центрувати

вихідні значення Y_i у формулах (4.5, 4.6). У позначенні вектора параметрів $\mathbf{a}_{kj}^{A}$ та окремого параметра a_{kjq}^{A} індекс *A* вказує, що параметри знаходяться на навчальній вибірці *A* об'єкта.

Розрахунок параметрів та порядок перебору структур моделей

Алгоритм COMBI знаходить найкращу структуру у складі повного полінома:

$$y = a_0^{'} + \sum_{i=1}^{M} a_i^{'} z_i + \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij}^{'} z_i z_j + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M} a_{ijk}^{'} z_i z_j z_k + \ldots, \quad (4.7)$$

де кількість складових сум у (4.7) дорівнює *p*, кількість членів полінома (4.7) $m = C_{M+p}^{M} - 1$, кількість підполіномів багаточлена (4.7) $q = 2^m - 1$. Структурні елементи повного поліному назвемо узагальненими змінними (УЗ) та позначимо як $x_i, i = 1,..., m$. У нових позначеннях вираз (4.7) матиме лінійний вигляд:

$$y = a_0' + \sum_{i=1}^{m} a_i' x_i \,. \quad (4.8)$$

Щоб знайти параметри деякого підполінома (4.9) з відповідним набором УЗ x потужності $\tilde{m}$

$$y_k = a_0 + a_1 x_{i_1} + \ldots + a_{\tilde{m}} x_{i_{\tilde{m}}} \quad (4.9)$$

необхідно сформувати інформаційну матрицю $(X^T X)_{\tilde{m}\times\tilde{m}}$ і вирішити нормальну систему рівнянь, отримавши рішення

$$\mathbf{a} = (X^T X)^{-1} X^T Y \,. \qquad (4.10)$$

Можливо організувати перебір підполіномів полінома (4.7) без оптимізації процесу: спочатку перебиратимуться всі лінійні структури, потім квадратичні і т.д. до поліномів степіню p. Кількість операцій (і час обчислень) у цьому підході для знаходження (4.10) для (4.9) буде пропорційним $\tilde{m}^3$. Однак можна зменшити обчислювальну складність на порядок - до $\tilde{m}^2$[5], організувавши послідовності вкладених структур і використавши рекурентні формули методу обрамлення в МНК [14] (формули РМНК) для обчислення параметрів структур, що перебираються. Обґрунтування застосування схеми перебору моделей з розрахунком параметрів рекурентним МНК у комбінаторному алгоритмі МГУА було запропоновано професором Степашко В.С. у роботі [33]. Для використання запропонованої процедури розрахунку параметрів необхідно організувати перебір структур таким чином, щоб конкретні структури для розрахунку параметрів вибиралися в порядку, що утворює комплекс послідовно вкладених структур (x_{i_1}), (x_{i_1}, x_{i_2}),.., $(x_{i_1}, x_{i_2}, ..., x_{i_k})$. Механізм утворення вкладених комплексів адитивних структур виду

$$Y_k = a_{d1} x_{i_1} ... + a_{d(k-1)} x_{i_{k-1}} + a_{dk} x_{i_k}, \quad d = 1,..., n \qquad (4.11)$$

видно з графічного представлення дерева повного перебору адитивних структур (рис. 4.6).

На рис. 4.6 представлений двійковий лічильник, при додаванні до останнього біта якого «1», генерується послідовний, праворуч наліво, шлях по дереву перебору. Лічильник встановлює відповідність між десятковими значеннями, як номерами структур, при обході дерева, та одиницями в ньому, як індикаторами (за розташуванням розряду) входження до структури відповідної УЗ.

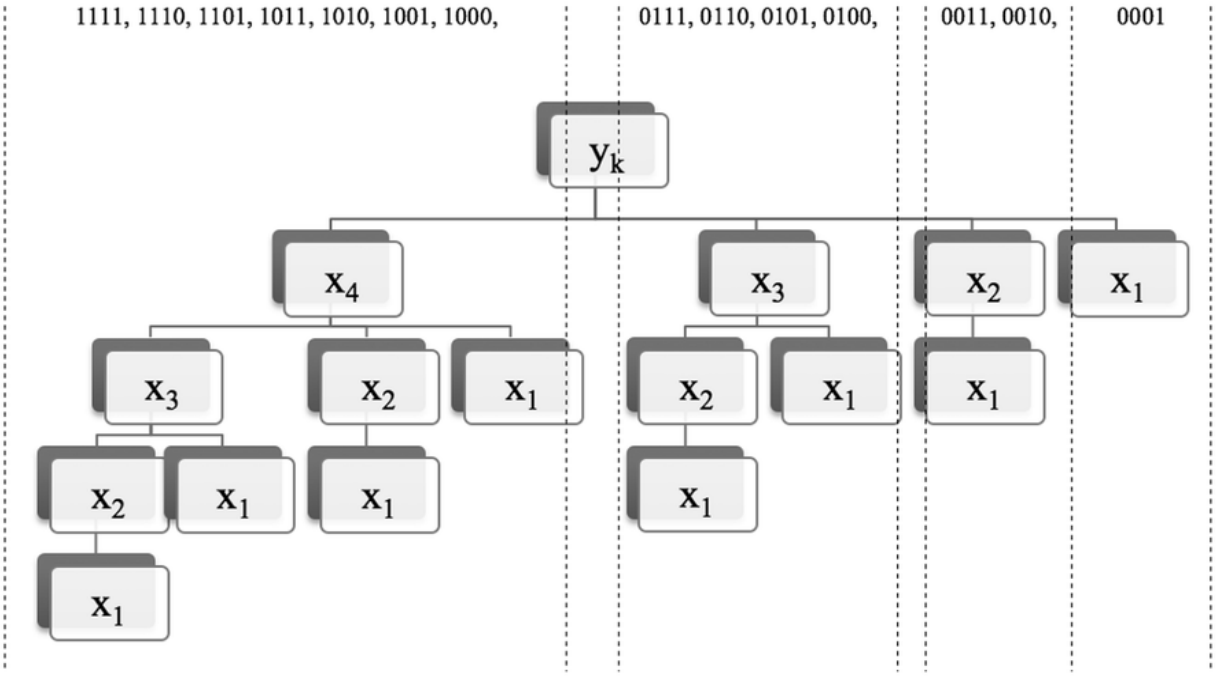


Рисунок 4.6. - Дерево повного перебору адитивних структур

Гілки дерева від кореня x_p до кінцевого листа x_1 утворюють вкладені структури. Розглянемо поточний стан лічильника на тій частині дерева, яка росте з кореня *p* на рівні *k* . Нехай сума «за модулем 2» поточного та наступних станів лічильника буде значенням, що складається лише з однієї одиниці у будь-якому з бітів. Тоді два зазначені стани лічильника відображають вкладені структури, а вершини дерева задовольняють умові суміжності і знаходяться на одній гільці на рівнях *k* та *k*+1. Якщо ця сума відрізняється від зазначеного значення, стан лічильника переміщує локацію на інші гілки дерева. Сума «по модулю 2», що представляє стан лічильника з двома одиницями у будь-яких розрядах, вказує на перехід на наступну гілку зліва на тому ж рівні *k* із спільним розгалуженням дерева для попереднього стану на рівні *k* -1. Сума «за модулем 2», що становить стан лічильника з трьома одиницями, вказує на перехід поточного стану дерева на наступну гілку зліва на рівні *k* -1 із спільним предком попереднього стану в цій гілці на рівні *k*-2. І так далі. Вказані властивості дозволяють легко визначити гілки вкладених структур дерева перебору та забувати інформацію у вузлах для рекурентного обчислення в момент підрахунку всіх нащадків кожного

предка.

Тепер можна використовувати рекурсивні формули РМНК для розрахунку параметрів моделі для адитивних вкладених структур, які відповідають гілкам дерева перебору. У кожній вершині дерева обчислюється:

1. Інформаційна матриця та її зворотня для кожного об'єкта *d*;
2. Вектор параметрів моделі $\mathbf{a}_d$ на заданій структурі на навчальних вибірках кожного об'єкта *d*;
3. Значення критерію селекції Cr_i що характеризує структуру;
4. Число *s*, що дорівнює кількості вже розрахованих нащадків для кожного предка.

Таким чином обчислюється кожна $i = 1,..., q = 2^m - 1$ вершина дерева. Обхід дерева завершується, коли стан лічильника досягає максимального значення. В результаті вибирається структура, на якій критерій вибору прийняв мінімальне значення. Наведемо реалізацію розрахунку параметрів рекурентним МНК для деякої гілки дерева перебору структур.

Припустимо, що на рівні *k* -1 дерева є деяка адитивна структура, $x_{i_1},..., x_{i_{k-1}}$ обчислені відповідні інформаційні матриці $G_{k-1} = (X^T X)_{(k-1)\times(k-1)}$ та зворотні їй $G_{k-1}^{-1} = (X^T X)^{-1}_{(k-1)\times(k-1)}$. Для структури $x_{i_1},..., x_{i_{k-1}}, x_{i_k}$ наступного рівня дерева перебору нам відома інформаційна матриця G_k наступного кроку. Рекурентний розрахунок параметрів $\mathbf{a} = G_k^{-1} X^T Y$ моделі $Y_k = a_1 x_1 ... + a_{k-1} x_{k-1} + a_k x_{i_k}$ алгоритмом COMBITOS GMDH вимагає застосування механізму РМНК для розрахунку матриці $G_k^{-1} = (X^T X)_{k\times k}$. Відповідний механізм розрахунку наведено у розділі 3.6.2. Для нерозривності викладу матеріалу повторимо його тут:

Нехай на кроці *k* відома інформаційна матриця

$$G_{k-1} = \begin{vmatrix} g_{1,1} & \cdots & g_{1,k-1} \\ \cdots & \cdots & \cdots \\ g_{k-1,1} & \cdots & g_{k-1,k-1} \end{vmatrix} \tag{4.12}$$

попереднього кроку розрахунку. Якщо матрицю наступного кроку G_k записати у блочному вигляді

$$G_k = \begin{vmatrix} G_{k-1} & u_k \\ v_k & g_{kk} \end{vmatrix}, \tag{4.13}$$

де $v_k = (g_{k,1},..., g_{k,k-1})$, $u_n = (g_{k,1},..., g_{k,k-1})^T$, то зворотна матриця G_k^{-1}, може бути записана як

$$G_k^{-1} = \begin{bmatrix} G_{k-1}^{-1} + \dfrac{G_{k-1}^{-1} u_k v_k G_{k-1}^{-1}}{g_k} & -\dfrac{G_{k-1}^{-1} u_k}{g_k} \\ \dfrac{v_k G_{k-1}^{-1}}{g_k} & \dfrac{1}{g_k} \end{bmatrix}, \tag{4.14}$$

де $g_k = g_{kk} - v_k G_{k-1}^{-1} u_k$. Слід зазначити можливість зменшення розмірності інформаційної матриці за рахунок розрахунків моделей у центрованих змінних

$$Y = y - \bar{y}_A, \; X_{i_k} = x_{i_k} - \bar{x}_{A,i_k}, \tag{4.15}$$

де індекс A позначає навчальну підвибірку, (далі індекс A опущений). В отриманих моделях вільний член дорівнюватиме нулю, значення критерію для цих моделей зміниться на константу, що дозволить знайти найкращі структури моделі в центрованому вигляді. Після визначення найкращої структури для центрованих моделей

$$Y_{\text{k}} = a_1^k x_{i_1} + ... + a_k^k x_{i_k} \tag{4.16}$$

можна класифікувати об'єкти у параметричному просторі центрованої версії змінних.

При необхідності можна знайти вільний член моделей, замінюючи центровані значення змінних через вихідні значення та їх середні, тоді отримаємо

$$(y-\bar{y}) = a_1^k (x_{i_1} - \bar{x}_{i_1}) + ... + a_k^k (x_{i_k} - \bar{x}_{i_k}) \quad (4.17)$$

Звідси з урахуванням

$$y = a_0^k + a_1^k x_{i_1} + ... + a_k^k x_{i_k} \quad (4.18)$$

знаходимо рівняння для

$$a_0^k = \bar{y} - a_1^k \bar{x}_{i_1} - ... - a_k^k \bar{x}_{i_k} \quad (4.19)$$

В результаті алгоритм визначення структури простору параметричних ознак завдання класифікації зводиться до наступних етапів:

1. Попередня обробка даних:

а). Формування матриці узагальнених змінних;

б). Перехід до матриці центрованих змінних.

2. Формування інформаційної матриці максимального розміру;

3. Організація перебору структур шляхом формування вкладених структур на дереві перебору. Для кожної згенерованої структури виконуються такі дії:

а) Формування інформаційної матриці заданої структури кожного об'єкта *d* з використанням результату п. 2;

б) Розрахунок параметрів моделі на отриманій структурі для кожного об'єкта *d* за допомогою РМНК;

с) Розрахунок значення зовнішнього критерію оцінюваної структури.

4. Вибір найкращої структури шляхом мінімізації значення зовнішнього критерію.

Використання результатів

З реалізації етапів 1-4 отримуємо оптимальну структуру $(x_{i_1}, x_{i_2}, ..., x_{i_k})$ і виконуємо перерахунок параметрів $\mathbf{a}_d$ для n моделей

$$Y_k = a_{d_1} x_{i_1} ... + a_{d_{(k-1)}} x_{i_{k-1}} + a_{d_k} x_{i_k}, d = 1,..., n \qquad (4.20)$$

по всій робочій вибірці W матриці центрованих змінних. Отримані вектори $\mathbf{a}_d$ утворюють матрицю ознак класів $k = 1,..., K$ об'єктів $d = 1,..., n$ у просторі параметрів A. Потім необхідно проаналізувати властивості класів у A, щоб вибрати відповідний алгоритм класифікації з постановки (*).

4.1.5.2. Ітераційний алгоритм перетворення простору спостережень за МГУА (OSCIA GMDH)

Основна відмінність OSCIA GMDH від попереднього алгоритму у тому, що він реалізує скорочений перебір структур, чим суттево розширює можливості підходу до вирішення завдань класифікації множин.

Моделі шукаються у класі структур

$$y_1 = a_0 + a_1 x_{i_1}, \ .... \ , \ y_k = a_0 + a_1 x_{i_1} + ... + a_k x_{i_k}, ... , \qquad (4.21)$$

де узагальнені аргументи x_{i_j} вибираються з повного набору поліноміальних членів (4.7) заданого степіню p.

Далі позначимо $z_j \quad j = 1,...,m$ - ирозширений набір вихідних змінних. Набір складений із перетворень

$$z'_j, (z'_j)^{-1}, (z'_j)^{1/3}, (z'_j)^{-1/3},\ j = 1,...,M \qquad (4.22)$$

вихідних змінних z'_j для можливості присутності у моделях аргументів, як і позитивних, так і у негативних раціональних степінях. На кожній наступній ітерації відповідно до значення зовнішнього критерію до структури додається найкраща узагальнена змінна у вигляді окремого поліноміального члена (4.7). Наведемо основні етапи алгоритму OSCIA GMDH :

1.Формування матриці розширеного набору вихідних ознак відповідно (4.22);

2.Розрахунок матриці узагальнених змінних *x*:

Кожна УЗ є одним із членів полінома (4.7) або одним з можливих множень змінних розширеної множини до степіня, обмеженого максимальним степінем полінома p . Якщо у вихідній розширеній матриці кількість змінних m, то найкращий спосіб генерувати УЗ полягає в тому, щоб сформувати послідовність чисел (p +1)-ої системи счислення без тих чисел, у яких сума цифр більша за p . Кількість чисел розраховується, як $S = C_{m+p}^{m} - 1$ - кількість членів у повному многочлені. Отримані S чисел в (p+1) системі счислення є позначення УЗ, де номер біта відповідає змінним з розширеної матриці, а значення біта демонструє степінь кожної розширеної змінної в поточній УЗ. Близький по суті підхід був використаний у роботі [33].

3. Розрахунок інформаційної матриці:

Щоб уникнути повторних обчислень при розрахунку параметрів моделей $X^T X$ на кожному етапі алгоритму розраховується інформаційна матриця максимального розміру $X^T X$ один раз, потім для формування $X^T X$

щоразу використовується тільки відповідні частини $X^T X$.

4. Генерація структур та оцінка параметрів:

Структура на ітерації k формується відповідно до виразу $y_k = a_0 + a_1 x_{i_1} + ... + a_k x_{i_k}$, де члени $x_{i_1}, ... x_{i_{k-1}}$ отримані попередніх етапах, а УЗ x_{i_k} визначається на поточному. Для пояснення механізму генерації УЗ на поточному етапі використовуємо дерево перебору (рис. 4.7). Перебір УЗ реалізується рухом по вузлах дерева зліва направо та зверху донизу.

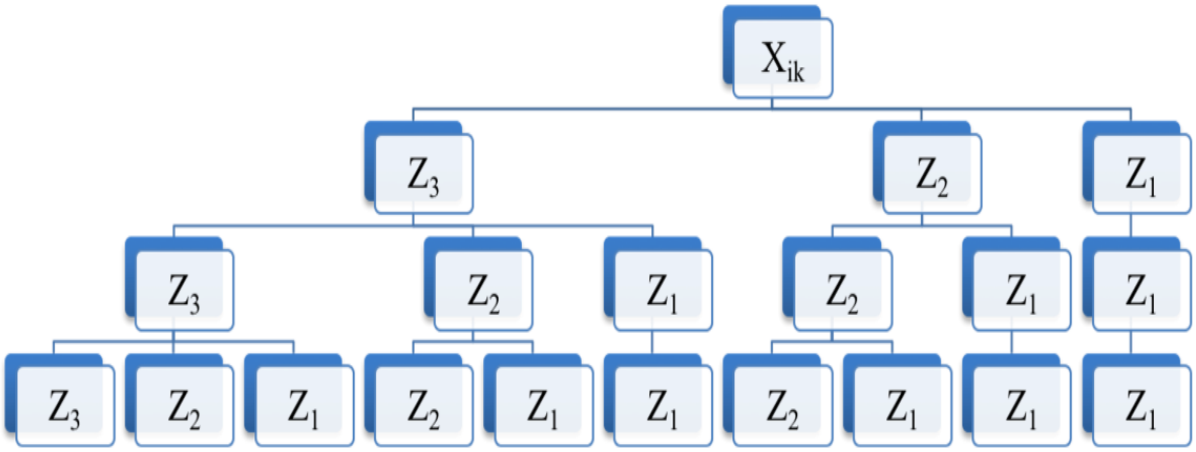


Рисунок 4.7. - Дерево перебору мультиплікативних членів

Одночасно створимо пов'язану з деревом таблицю поліноміальних членів (табл. 4.1). Саме її образ буде використано для реалізації перебору в алгоритмі. Кожен рядок таблиці відображає деяку УЗ, цифра в клітині таблиці - ступінь відповідної змінної z, що входить до УЗ. Переміщення по дереву перебору зліва направо і зверху вниз відповідає підйому по таблиці з знизу вгору, спочатку рядками, сума значень яких дорівнює 1, потім, рядками, де сума значень дорівнює 2 і т.д. до (p +1).

Таблиця 4.1

Схема перебору поліноміальних членів

Z_3	Z_2	Z_1
0	0	1
0	0	2
0	0	3
0	1	0
0	1	1
0	1	2
0	2	0
0	2	1
0	3	0
1	0	0
1	0	1
1	0	2
1	1	0
1	1	1
1	2	0
2	0	0
2	0	1
2	1	0
3	0	0

В алгоритмі використовується два параметри, що керують кількістю структур, що відбираються на етапах алгоритму: F_1 і F_2. Значення параметра F_1 використовується на першому етапі, F_2 на інших. Крім того, задається параметр F - кількість кращих структур, які зберігаються, як результат роботи всього алгоритму загалом. Оскільки моделі $y_k = a_0 + a_1 x_{i_1} + ... + a_k x_{i_k}$ формуються шляхом послідовного додавання нових узагальнених змінних, вони утворюють комплекси вкладених структур. Тому для розрахунку параметрів моделей може бути використаний рекурентний МНК аналогічно версії COMBIT OS GMDH (див. розділ 4.1.5.1).

5. Зовнішній критерій алгоритму.

У зв'язку з організацією в алгоритмі OSCIA скороченого перебору структур, тут можна пропонувати складніші, але й ефективніші зовнішні

критерії:

а) Комбінований критерій точності:

$$Cr_{sel} = (1-\beta)\cdot(\alpha\cdot\Delta_A^{\gamma} + (1-\alpha)\cdot\Delta_B^{\gamma}) + \beta\cdot\left|\frac{\Delta_B^{\gamma} - \Delta_A^{\gamma}}{\Delta_B^{\gamma} + \Delta_A^{\gamma}}\right| \quad (4.23)$$

де значення α, γ, Δ_A^{γ}, Δ_B^{γ} вказані в (4.5), (4.6) розділу 4.1.5.1. Нова адитивна складова критерію (4.23) є критерій збалансованості, тут β - вага критерію збалансованості в комбінованому критерії.

б) Критерій роздільності .

Оскільки кінцевою метою нашої задачі є класифікація, логічно оцінювати якість шуканої структури з погляду на її здатність розділяти класи. Тому пропонується обчислювати зовнішній критерій як відношення внутрішньокласової дисперсії до міжкласової в просторі параметрів моделей об'єктів класифікації. Тут під міжкласової дисперсією розуміється розкид центрів у парах класів.

$$Cr_{sep} = \frac{\frac{1}{K}\sum_{j=1}^{K}\frac{1}{n_j}\sum_{r_i\in R_j}^{n_j}\left\|\mathbf{a}_i - \bar{\mathbf{a}}_j\right\|^2}{\frac{1}{C_K^2}\sum_{i=1}^{K}\sum_{j=i+1}^{K}\left\|\bar{\mathbf{a}}_i - \bar{\mathbf{a}}_j\right\|^2}, \quad (4.24)$$

де C_K^2 - кількість комбінацій з K по 2.

З урахуванням дисперсії за кожною з m координат ($p = 1,..., m$, де $m-$ розмірність простору $A = R^m$) та розуміння норми як евклідової відстані (4.25) вираз (4.24) набуває вигляду (4.26). Для зручності запису поточний індекс класу q був позначений, як верхній індекс:

$$\left\|\mathbf{a}_i - \overline{\mathbf{a}}_j\right\|^2 = \sum_{p=1}^{m}(\mathbf{a}_{ip} - \overline{\mathbf{a}}_p^q)^2; \tag{4.25}$$

$$Cr_{sep} = \frac{\frac{1}{K}\sum_{q=1}^{K}\frac{1}{n_q}\sum_{i=1}^{n_q}\sum_{p=1}^{m}(a_{ip} - \overline{a}_p^q)^2}{\frac{1}{C_K^2}\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{p=1}^{m}(\overline{a}_{ip} - \overline{a}_p^q)^2} \tag{4.26}$$

4.1.6. Алгоритм пошуку найкращих структур кожного класу

Вочевидь, що специфіка класів може вимагати описати об'єкти, що належать різним класам, різними структурами. Це ініціює необхідність пропонувати і нові алгоритмів отримання оптимальних структур вже для кожного класу і способи їх застосування для класифікації. Нижче пропонується скоригована версія алгоритму OSCIA, яка шукає структури, що найбільш підходять для відділення об'єктів кожного класу від об'єктів усіх інших класів у просторі параметрів їх моделей. Тут оцінюється якість кожної структури у кожному класі.

Як бачите, обидва алгоритми мають багато спільного. Основна модифікація стосується зміни зовнішніх критеріїв. Вони розраховують компоненти помилки $E(k)$ та дисперсії $D(k)$ у просторі параметрів моделей окремо для кожного класу k:

$$Cr(k) = \beta E(k) + (1-\beta)D(k), \tag{4.27}$$

де β визначає співвідношення ваг компонентів критерію, $E(k)$ розраховується по похибкам для поточного класу k та класу, зведеного з інших класів k':

$$E(k) = \left(\frac{Err(k) + Err(k')}{\sum_{i \in A} (Y_i)^2}\right)^{1/2} \tag{4.28}$$

Позначення $Err(k)$ і $Err(k')$ мають на увазі, що:

$$Err(k) = \gamma_k \sum_{j}^{n_k} \sum_{i \in A(k,j)} \left(Y_{kji} - \sum_{p}^{M} x_{ip} a_{kjp}^{A} \right)^2 \tag{4.29}$$

$$Err(k') = \gamma_k' \left(\sum_{q=1}^{k-1} \sum_{j}^{n_q} \sum_{i \in A(q,j)} \left(Y_{qji} - \sum_{p}^{M} x_{ip} a_{qjp}^{A} \right)^2 + \sum_{q=k-1}^{K} \sum_{j}^{n_q} \sum_{i \in A(q,j)} \left(Y_{qji} - \sum_{p}^{M} x_{ip} a_{qjp}^{A} \right)^2 \right. \tag{4.30}$$

Коефіцієнти класів γ_k та γ_k' можуть бути задані рівними 1 і 0, інші значення дозволяють досліджувати можливості алгоритму. Визначення оптимальних структур S_i°, $i = 1,\ldots,K$ коли $\beta = 1$ передбачає використання критерієм точність представлення об'єктів моделлю.

Дисперсійна частина критерію (15) характеризує властивості структури відокремлювати «свій» класвід інших класів. Він заснований на мінімізації дисперсії у класах та максимізації відстані між центрами класів:

$$D(k) = \frac{\alpha \cdot D^{in}(k) + (1-\alpha) \cdot D^{in}(k')}{Dist(k)}, \tag{4.31}$$

де α дозволяє, за необхідності, варіювати перевагу дисперсії об'єктів «свого» класу щодо структури. Дисперсія всередині класу k та інших класів k':

$$D^{in}(k) = \frac{1}{n_k} \sum_{i=1}^{n_k} \sum_{p=1}^{m} \left(a_{ip}^{k} - \bar{a}_{p}^{k} \right)^2 , \tag{4.32}$$

$$D^{in}(k') = \sum_{j=1}^{k-1} \frac{1}{n_q} \sum_{i=1}^{n_q} \sum_{p=1}^{m} \left(a_{ip}^{q} - \bar{a}_{p}^{q} \right)^2 + \sum_{j=k+1}^{K} \frac{1}{n_q} \sum_{i=1}^{n_q} \sum_{p=1}^{m} \left(a_{ip}^{q} - \bar{a}_{p}^{q} \right)^2 \tag{4.33}$$

Відстань між центром класу k та центром класу k' (об'єднання всіх точок, крім тих, що належать класу k) дорівнює:

$$Dist(k) = \sum_{p=1}^{m} \left(\bar{a}_{p}^{k} - \bar{a}_{p}^{k'} \right)^2 , \tag{4.34}$$

де $\bar{a}_{p}^{k'}$ є центром класу k' . В результаті отримуємо стільки структур, скільки класів, причому кожна структура є найбільш релевантною для відповідного класу. Визначення оптимальних структур S_i° $i = 1,\ldots,K$, при $\beta = 0$ передбачає розробку системи класифікації використовує принцип «один проти всіх» у просторі отриманих структур .

4.1.7. Рекомендації до застосування та перспективи підходу

Визначимо перспективи розвитку підходу.

1. Розглянутий підхід очевидно може бути застосованим до завдання класифікації зображень. Для цього потрібно запропонувати набір характеристик зображення, зміна залежності між якими від зображення до зображення зможе стати основою для їх класифікації. Крім того, підхід покаже нові можливості під час використання широкого спектру різних зовнішніх критеріїв, що дозволить формувати структури з найбільшою ефективністю для вирішення завдань класифікації.

2. Умовою застосування запропонованої технології є припущення існування для об'єктів класифікації статистичних залежностей у явному

вигляді. Надалі може передбачатися більш загальний підхід, що враховує зв'язки між змінними у неявній формі, що вимагатиме запропонувати відповідні алгоритмічні рішення .

3. Можна окреслити перспективність запропонованого підходу як інструменту для вирішення завдань «великих даних», не обмежуючись проблемою КМ. Пропоновані алгоритми пошуку оптимальних структур і моделей несуть у собі сенс цілеспрямованої згортки вихідних множин у стислу форму із збереженням цільової інформації (яка визначається зовнішнім критерієм) у структурі та параметрах моделей об'єктів. Зазначимо, що цей інструмент представляє самостійний інтерес, коли можна розглядати для зберігання, передачі та цільового застосування не вихідні множини спостережень, а моделі, побудовані на їх основі. Окремо можуть розглядатися умови за яких може бути вирішена і зворотна (в загальному випадку некоректна) задача – відновлення вихідних множин за відомими їх моделями.

4.1.8. Кластеризація множин.

Формально завдання класифікації множин містить, як завдання з учителем, так і завдання без вчителя. Намітимо нижче один із можливих підходів до кластеризації об'єктів, заданих множинами спостережень - перехід для вирішення задач кластерного аналізу до параметричного простору моделей об'єктів.

Як і раніше вважатимемо, що кожен об'єкт кластеризації задається деяким набором спостережень, виміряних при різних станах середовища. Допускається частковий перетин областей, що становлять об'єкти кластеризації у просторі ознак. Основними припущеннями для застосування однієї з версій розробленого підходу є три:

1. Наявність статистичного зв'язку між ознаками, що характеризують об'єкти.

2. Наявність єдиної структури (назвемо її ядром кластеризації), з урахуванням якої можливе отримання досить точних описів об'єктів

кластеризації.

3. Відмінності у параметричному представленні об'єктів дають очікуваний ефект їх кластеризації.

Умови 2 і 3, як очевидно, не є чимось суворим, формальним, оскільки не визначено, що вважати «досить точними описами» та «очікуваним ефектом». Однак сама постановка завдання кластеризації, за своєю суттю, не дає однозначного рішення, і будь-яка кластеризація піддається досить складному етапу інтерпретації результатів, що є вирішальним аргументом в оцінці корисності чи слабкості застосування підходу. При застосуванні інструментів кластерного аналізу завжди доводиться покладатися на значну роль та інтуїцію оператора аналізу, евристики є одним з найважливіших компонентів відповідних алгоритмів.

У цих припущеннях кластеризація об'єктів може бути виконана з використанням наступного двоетапного механізму:

1.Вирішується проблема знаходження загальної структури для адекватного опису всіх об'єктів завдання кластеризації. Застосуємо описаний вище варіант алгоритму моделювання з допомогою зовнішнього критерію точності. За результатами рішення одержуємо параметри моделі кожного об'єкта. Значення параметрів будуть значеннями ознак у новому просторі ознак задач кластеризації.

2. На розсуд оператора-дослідника виконується кластеризація об'єктів будь-яким із відповідних аналізу методів, дається інтерпретація отриманих кластерів та оцінюється результат кластеризації.

Окремо може бути розглянуто випадок, коли дані для кластеризації отримуються в результаті активного експерименту. У цьому випадку кількість умов експерименту для кожного спостереження об'єкта може задаватися однаковим, рівним m , і кожне із спостережень записується за відповідних однакових умов. Тоді новий простір формується з початкових змінних, але шляхом їхнього повторення, кратного m, у відповідних експериментальних умовах. Таким чином, нова розмірність результуючого

простору кластеризації об'єктів дорівнює $m * n$, де n - розмірність вихідного простору. Інформація про об'єкт у цьому просторі (багатомірні спостереження щодо об'єкта) зосереджена в одній багатовимірній точці в розширеному просторі розмірності $m*n$. Тому кластеризація може бути реалізована звичайними методами кластерного аналізу.

4.1.9. Контрольні запитання

1. Наведіть та обгрунтуйте методи класифікації множин при незалежних вхідних ознаках

2.Обгрунтуйте ідею підходу для класифікації множин через єдину спільну структуру моделей об’єктів класифікації

3. Наведіть та обгрунтуйте комбінаторний алгоритм МГУА для пошуку єдиної спільної структури об’єктів класифікації

4. Наведіть та обгрунтуйте ітераційний алгоритм МГУА для пошуку єдиної спільної структури об’єктів класифікації

5. Обгрунтуйте застосування критерію роздільності при застосуванні у алгоритмах пошуку моделей об’єктів класифікації

6.Обгрунтуйте підходи для класифікації множин при різних оптимальних структурах моделей об’єктів у класах

7. Обгрунтуйте можливості розглянутого підходу для кластеризації об’єктів заданих множинами спостережень

8. Опишить перспективи підходу для класифікації зображень

4.2. Класифікація динамічних об’єктів за особливостями причинно-наслідкової структури

4.2.1 Передумови для побудови алгоритму

Розглядається задача класифікації, в якій объекти характеризуються не поодинокими спостереженнями у багатовимірному просторі ознак, а їх множинами, що є реалізаціями часових рядів. Для спрощення, далі пропонується обмежити формалізацію випадком для двух классов. Об’єкти A_i та B_j класів A і B характеризуються реалізациями $x_{pj}, j = 1, \ldots,$ n часових

рядів (процесів) X_p, $p=1,..,m$.

Ряди ознак X_p, безпосередньо незручно застосовувати для оцінки міри близькості об'єктів до певного класу. Тому далі запропонуємо деяку процедуру формування вторинних ознак, що дозволяє застосувати звичні міри близькості та алгоритми класифікації [34]. Дана процедура має використовувати доцільні критерії та параметри селекції для отримання найкращої якості результатів класифікації. Ефективність запропонованого нижче підходу залежить від ступеня виконання наступних припущень:

1. Часові ряди X_p, $p=1,..,m$, що характеризують об'єкт є взаємопов'язані процеси.

2. Взаємозв'язок процесів X_p, $p=1,..,m$ значною мірою має характеризуватись лінійним ефектом.

Логічно вважати, що при виконанні припущень методу однією з характерних ознак об'єктів, що класифікуються, можуть бути статистичні причинно-наслідкові структури (ЧНС) процесів X_p, $p=1,..,m$. А відмінності у ЧНС об'єктів різних класів у певних випадках можуть бути основою для класифікації.

Як ми розглядали у розділі 2 визначення структури ЧНЗ може бути отримано за допомогою розрахунків ККФ, чи застосуванням процедур визначення причинності по Грейнджеру, чи методик статистичного причинно-наслідкового аналізу. В результаті застосування конкретного критерію при конкретних значеннях порогів отримаємо для кожного об’єкту матрицю ЧНС, що відповідає даному вибору. У припущенні, що побудовано матриці для кожного об’єкту визначимо далі критерії та параметри селекції ЧНС, що мають дати найкращі результати класифікації за порівнянням ЧНС об’єктів класифікації.

4.2.2. Міри близкості та можливі підходи до класифікації

Нехай матриці ЧНС A_k об'єктів класу А і матриці ЧНС B_k об'єктів класу В мають вигляд:

$$A_k = \begin{matrix} a_{11}^k & \dots & a_{1m}^k \\ \dots & \dots & \dots \\ a_{m1}^k & \dots & a_{mm}^k \end{matrix}, \; B_k = \begin{matrix} b_{11}^k & \dots & b_{1m}^k \\ \dots & \dots & \dots \\ b_{m1}^k & \dots & b_{mm}^k \end{matrix}. \qquad (4.35)$$

Враховуючи припущення про ЧНС, як характеристику класів об'єктів, заданими часовими рядами, необхідно вибрати критерій Cr_i і значення порогів селекції ЧНС, що забезпечують найближчі матриці ЧНС на об'єктах одного класу (відповідно до мінімуму внутрішньокласової дисперсії відповідних елементів матриць):

$$Cr_{cl-} = \sum_g^{N_A} \sum_h^{N_A} (\sum_j^m \sum_i^m (a_{ij}^g - a_{ij}^h)^2 + \sum_g^{N_B} \sum_h^{N_B} (\sum_j^m \sum_i^m (b_{ij}^g - b_{ij}^h)^2 \qquad (4.36)$$

і найбільш відмінні матриці ЧНС на об'єктах різних класів (відповідно до максимуму міжкласової дисперсії відповідних елементів матриць):

$$Cr_{cl+} = \sum_g^{N_A} \sum_h^{N_b} (\sum_j^m \sum_i^m (a_{ij}^g - b_{ij}^h)^2. \qquad (4.37)$$

Тоді вибір конкретного критерію та порогів селекції ЧНС для завдання класифікації можна розраховувати відповідно до таких дисперсійних критеріїв:

$$\mathrm{C}r_{cl} = \max_{Cr_i, \Delta_1, \Delta_2} (\frac{\alpha}{Cr_{cl-}} + \beta \cdot Cr_{cl+}), \qquad (4.38)$$

де коефіцієнти α та β балансують вимоги збігу та розрізнення матриць ПСС або аналог критерію роздільності класів:

$$\mathrm{C}r_W = \max_{Cr_i, \Delta_1, \Delta_2} \frac{Cr_{cl+}}{Cr_{cl-}}. \qquad (4.39)$$

Процедура максимізації критеріїв може враховувати розбиття вибірки

об'єктів класифікації: розрахунок параметрів для всіх варіантів критеріїв селекції на навчальній вибірці та вибір найкращого варіанта комплекту по максимуму дисперсійного критерію (4.38) або (4.39) на перевірочній вибірці .

Міру близькості об'єкта до класу, можна прийняти, як суму відмінностей матриці ЧНС класифікованого об'єкта від матриць ЧНС об'єктів кожного класу. Природні алгоритми класифікації: визначення найкращого граничного значення міри близькості, що розділяє об'єкти класів (дискримінантний аналіз), алгоритм найближчого сусіда, чи алгоритм зваженого пооб'єктного голосування у кожному класі.

4.2.3. Контрольні запитання

1. Назвіть передумови для застосування результатів причинно-наслідкового аналізу для класифікації динамічних об’єктів.

2. Обгрунтуйте за яким критерієм оцінити близкість ЧНС об’єктів одного класу

3. Обгрунтуйте за яким критерієм оцінити відмінність ЧНС об’єктів різних класів

4. Обгрунтуйте комплексні критерії для оцінки параметрів ЧНА для забезпечення якісної класификації динамічних об’єктів.

4.3 Застосування методів математичного програмування для вирішення задач моделювання.

4.3.1 Передумови для застосування АМП для моделювання

Сучасне моделювання (в тому числі аналітичне) прийнято відслідковувати від групи вчених пов'язаних з Ньютоном, який створив теорію нескінченно малих - теорію диференціального і інтегрального обчислення. Це дивовижне прозріння практично вперше породило аналітичні математичні моделі в точності, відповідні реальним фізичним процесам.

Ньютон висунув геніальну по простоті і адекватності ідею, що значна кількість фізичних моделей макросвіту пов'язують у своїх взаєминах лінійні переміщення з їх швидкостями, прискореннями і т.д.. і створив адекватний

апарат опису об’єктів з даними змінними. За допомогою цієї теорії було вирішено грандіозне завдання - завдання моделювання руху тіл небесної механіки. Паралельно ця задача породила до життя супутні напрямки досліджень - теорію ймовірності (роботи Лагранжа, Муавра, Байєса). Звичайно теорія ймовірності має свої власні корені, пов'язані з вивченням випадковості, але обчислювальні, практичні аспекти теорії ймовірності стимулювало споріднене «небесній механиці» завдання з теорії вимірювань. Саме нижче зазначена задача стала першим сформульованим завданням статистичного моделювання.

Необхідно було зуміти найбільш точно визначати початкові умови для вирішення завдань Небесної механіки - тобто визначити найбільш точне положення зірок на небосхилі за неодноразовими вимірами цього положення.

Таким чином в 1800-тих роках Лаплас, Гаусс і Лежандр, кожен з яких в той час працював над теорією руху небесних тіл, і не міг оминути проблеми теорії вимірювань, запропонували свої варіанти вирішення такого завдання. Гауссом і Лежандром було запропоновано оцінювати невідоме значення вимірюваної величини a по його повторним вимірам $x_1, x_2, ..., x_n$, як таку величину $\hat{a} = a(x_1, x_2, ..., x_n)$, яка забезпечує мінімум функціоналу

$$\Delta = \min_{\hat{a}} \sum_{i=1}^{n} (x_i - \hat{a})^2 \qquad (4.40)$$

Це була перша постановка для розрахунку параметра методом найменших квадратів. Як ми знаємо рішення такої однопараметричної задачі буде середнє. У той же час Лаплапсом був запропонований для тих же умов задачі інший функціонал

$$\Delta = \min_{\hat{a}} \sum_{i=1}^{n} \left| x_i - \hat{a} \right| \qquad (4.41)$$

Виявилося, що таке значення $\hat{a}$ відповідає знаходженню вибіркової емпіричної медіани, тобто такому числу $\hat{a}$

$$\hat{a} = \begin{Bmatrix} \frac{1}{2}(x_k + x_{k+1}), n = 2k \\ x_{k+1} \qquad\qquad n = 2k+1 \end{Bmatrix}, \tag{4.42}$$

праворуч і ліворуч від якого знаходиться однакова кількість вимірювань.

Пізніше, було показано, що завдання мінімізації суми модулів відхилень може бути вирішено за допомогою задачі лінійного програмування (ЛП задачі). Покажемо, як можно сконструювати таку задачу. Спершу розглянемо спільне та відмінності, що демострують 2 відомі оптимізаційні підходи – МНК та ЛП задачі.

4.3.2. Спільне та відмінності в задачах ЛП та МНК

Нагадаємо, матричний запис канонічної ЛП задачі

$$\min_x c^T x \tag{4.43}$$

при обмеженнях $$Ax = b \quad x \geq 0 \tag{4.44}$$

або, як розписати (4.44), маємо

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{4.45}$$

Тут маємо m обмежень та n змінних x, при $c \in R^n$, $x \in R^n$, $b \in R^m$,

A = A(m,n), rankA =m, (m<n), $x = (x_1, x_2, ..., x_n)$, $\forall x_i \geq 0$.

Констатуємо спільні риси МНК та ЛП:

1. І МНК і ЛП - різновиди задач оптимізації;

2. І МНК і ЛП застосовуються для вирішення завдань моделювання, але в разі ЛП - ми можемо сформулювати більше різноманітних задач моделювання, так як існує природний елемент задач ЛП - можливо застосувати обмеження на область моделювання.

Відмінності:

- Оптимізаційна постановка МНК визначає функціонал завдання, як

$$\min \sum_i^n e_i^2 \quad , \tag{4.46}$$

де e_i - складові вектора нев'язок $E = XA - y$ умовної системи рівнянь

$$XA \cong y \quad \text{чи} \quad \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = * = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \tag{4.47}$$

Ця система визначає перевизначену систему рівнянь, так як в МНК постулюється умова, що число обмежень-вимог m (точок х) до рішення *більше* числа n – кількості змінних системи a_n (тобто розмірності вектора А)

Тиким чином, ми маємо в постановці МНК таку специфічну форму задачі оптимізації, коли у нас немає допустимих рішень, і ми знаходимо *найкраще рішення з недопустимих.*

А коли в задачі ЛП $m_{\text{обмежень}} > n_{\text{змінних}}$, то знайти рішення не можна, так як немає допустимої області рішень. Тобто рішення задачі оптимізації в формі задачі ЛП може бути застосовано тільки при недовизначеній системі типу (4.44), (4.45), тобто коли $m_{\text{обмежень}} < n_{\text{змінних}}$. Тоді для ЛП задачі існує область допустимих рішень.

4.3.3. Підходи до вирішення несумісних оптимізаційних задач

Повернемося знов до розгляду задачі (4.43), (4.44) і нехай, при цьому, маємо ситуацію $m_{\text{обмежень}} > n_{\text{змінних}}$. Спроба вирішувати її через задачу ЛП в лоб впирається в несумісність (4.44) і відсутність ОДР так як $m > n$.

Тоді ввівши з (4.44) вектор нев'язки:

$$y = Ax - b \tag{4.48}$$

можемо пропонувати 2 шляхи вирішення проблеми:

1. Варіант типу МНК**:**

Метод безумовної оптимізації при несумісній системі обмежень

Сформуємо критерій $L = c^T x + \alpha^T y^2$ і будемо вимагати його мінімізації

$$\min_{x,y} c^T x + \alpha^T y^2 \tag{4.49}$$

тут застосовуємо "y^2", так як невязка може бути, як додатнью, так і від'ємною. Підставивши у (4.49) нев'язку (4.48) маємо

$$\min_{x} L = \min_{x} c^T x + \alpha^T (Ax - b)^2 \tag{4.50}$$

Далі подібно до МНК беремо похідні до L і, прирівнюючи їх нулю, одержуємо систему лінійних рівнянь щодо x

$$\frac{\partial L(x)}{\partial x_j} = 0 \quad j = 1.,,,.n \tag{4.51}$$

Залишається питання балансування вимог до нашої задачі – складових

критерію: первинного фукціоналу $c^T x$ та нев'зки $Ax - b$**.** Це питання вибору вектора параметрів α , в тривіальному випадку беремо всі $\alpha = 1$.

4.3.4. Застосування задач ЛП при несумісній системі обмежень

Знов повернемося до розгляду задачі (4.43), (4.44) і нехай при цьому знов маємо ситуацію $m_{\text{обмежень}} > n_{\text{змінних}}$. Як ми відмічали вище, при такому випадку ОДР у ЛП задачі не існує, так як (4.44) не можуть бути задовільнені. Але, якщо розглянути мінімізацію вхідного функціоналу $c^T x$ и модулю вектору нев'язок $y = |Ax - b|$ тоді можливо одержати задачу вже в класі вирішуємих ЛП задач. Тут маємо на увазі наступне: розглянемо задачу

$$\min_{x.y} [c^T x + \alpha^T \cdot y] \tag{4.52}$$

$$Ax - \mathrm{b} - y \le 0 \tag{4.53}$$

$$Ax - \mathrm{b} + y \ge 0 \tag{4.54}$$

$$x \ge 0 \quad y \ge 0 \tag{4.55}$$

Звернемо увагу: задача (4.52)-(4.55) може бути еквівалентно приведена до канонічної форми

$$\begin{aligned} &\min_{x.y} [c^T x + \alpha^T \cdot y] \\ &Ax - \mathrm{b} - y + \mathrm{z}_1 = 0 \\ &Ax - \mathrm{b} + y - \mathrm{z}_2 = 0 \\ &x \ge 0 \, , y \ge 0, \mathrm{z}_1 \ge 0, z_2 \ge 0 \end{aligned} \tag{4.56}$$

де вже кількість $m_{обмежень} < n_{змінних}$, тобто тут ОДР формально існує.

Тепер повернемося до (4.52)-(4.55) і розмірковуємо:

Нехай одна із складових векторної нерівності (4.53) на деякому кроці алгоритму ЛП задачі прийме значення $Ax - b \ge 0$**.** Тоді, щоб виконувалося відповідне обмеження типу (4.53) відповідний "y" зобов'язаний бути деяким додатнім (не від'ємним) і таким, при цьому, щоб $y \ge Ax - b$, - тільки в цьому

випадку (4.53) буде виконуватися. При цьому відмічаємо, що наша задача ЛП працює на мінімізацію нев'язки "*y*" у нерівності (4.53), при цьому відповідна нерівність (4.54) виконується автоматично.

Тепер, нехай складова нерівності (4.54) прийме значення $Ax - b \le 0$. Тоді, щоб виконувалося відповідне обмеження типу (4.54) відповідний "*y*" зобов'язаний бути знов деяким додатнім (не від'ємним), і таким що, $y \ge |Ax - b|$, при цьому відповідна нерівність (4.53) виконується автоматично. Таким чином, в задачі (4.52)-(4.55) мінімізуємо *модуль невязки "y",* тому що структура нерівностей системи, як бачимо, завжди забезпечує $y \ge 0$. Вирішуючи (4.52)-(4.55) осягаємо 2 мети: здійснюємо мінімізацію вхідного функціонаалу $c^T x$ і забезпечуємо мінімізацію модулю вектора нев'язок *y*. Залишається, як і в першому випадку, необхідність балансувати вхідний критерій $c^T x$ з вимогою мінімізації нев'язки *y*. Цей компроміс є питання вибору вектора коефіцієнтів α^T. У тривіальному випадку візьмемо $\alpha = 1$. Далі наведемо окремий випадок вище наведеної задачі.

4.3.5. Завдання моделювання з мінімізацією модульного критерію невязок

Сформулюємо завдання апроксімації даних, виходячи з критерію мінімуму суми модулів невязок виходу та моделі.

Нехай маємо вхідні дані

Y	X_1	X_2	...	X_m
y_1	x_{11}	x_{12}	...	x_{1m}
y_2	x_{21}	x_{22}	...	x_{2m}
...	...	...	...	...
y_n	x_{n1}	x_{n2}	,,,	x_{nm}

(4.57)

для знаходження параметрыв $r_0, r_1, ..., r_m$ моделі $y = r_0 + \sum_{i=1}^{m} r_i x_i$ за критерієм мінімуму суми модулів невязок моделі. Формально таку задачу

можемо записати наступним чином:

$$\begin{cases} L = \min\limits_{r,s} \sum\limits_{j=1}^{n} \left| S_j \right| \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) = S_j, \quad j = 1,...,n \end{cases} \tag{4.58}$$

Як очевидно, дану задачу можно записати, як окремий випадок попереднього завдання (4.52)-(4.55):

$$\begin{cases} L = \min\limits_{r,s} \sum\limits_{j=1}^{n} S_j \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \le S_j, \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \ge -S_j, \; j = 1,...,n \\ S_j \ge 0 \end{cases} \tag{4.59}$$

Дану конструкцію двійних нерівностей доцільно застосовувати і у задачах оптимізації, де критерій формується не на мінімум чи максимум критеріальної змінної, а як мінімум її відхилення від заданого значення.

4.3.6 Завдання моделювання з одностороннім критерієм та критерієм мінімакс.

Тепер можемо легко записати цілу серію нових конструкцій для різних доцільних задач моделювання:

Задача моделювання з умови мінімума максимального відхилення моделі від табличних точок:

$$\begin{cases} L = \min\limits_{r,s} S \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \le S, \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \ge -S,\ j = 1,...,n \\ S \ge 0 \end{cases} \tag{4.60}$$

Задача для односторонньої мінімізації:

$$\begin{cases} L = \min\limits_{r,s} \sum\limits_{j=1}^{n} S_j \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \le 0, \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \ge -S_j,\ j = 1,...,n \\ S_j \ge 0 \end{cases} \tag{4.61}$$

чи

$$\begin{cases} L = \min\limits_{r,s} \sum\limits_{j=1}^{n} S_j \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \le S_j, \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \ge 0,\ j = 1,...,n \\ S_j \ge 0 \end{cases} \tag{4.62}$$

Задача для одностороннього мінімаксу:

$$\begin{cases} L = \min\limits_{r,s} \sum\limits_{j=1}^{n} S \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \le 0, \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \ge -S,\ j = 1,\ldots,n \\ S \ge 0 \end{cases} \tag{4.63}$$

чи

$$\begin{cases} L = \min\limits_{r,s} \sum\limits_{j=1}^{n} S \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \le S, \\ y_j - (r_0 + \sum\limits_{i=1}^{m} r_i \cdot x_{ij}) \ge 0,\ j = 1,\ldots,n \\ S \ge 0 \end{cases} \tag{4.64}$$

Відмітимо у завершенні, що вище не застосовані обмеження на додатність параметрів моделі, так як оминути це обмеження завжди можливо очевидною заміною $r_i = u_i - v_i$, де вже $u_i \ge 0$, $v_i \ge 0$.

4.3.7. Контрольні запитання

1. Сформулюйте задачу, що виникла, як передумова для застосування апарату МП для моделювання

2.Порівняйте методи оптимізації МНК та ЛП для моделювання, та сформулюйте області застосування

3. Наведіть та обгрунтуйте ЛП задачу з критерієм модулю відхилень

4.Наведіть та обгрунтуйте конструкції ЛП задач для моделювання з односторонніми критеріями та критерієм мінімаксу.

СПИСОК ЛІТЕРАТУРИ

1. Пасічник В.В., Виклюк Я.І., Камінський Р.М. Моделювання складних систем. Посібник. Львів: Видавництво "Новий Світ - 2000". 2017. 404 с.
2. Павлов В.А. Причинно-следственный анализ в системах процессов [Текст] / В.А. Павлов // Проблеми информаційних технологій. – 2009. - №5. - С. 8-14.
3. Granger C.W.J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods // Econometrica. 1969. No37(3). P. 424-438
4. Кондрашова Н. В. Метод Грама-Шмидта-Гаусса для решения систем линейных уравнений в МГУА / Кондрашова Н. В. // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. – 2009. – № 50. – С. 158–163. 5-142
5. Ivakhnenko, A., & Stepashko, V. (1985). Noise-immunity modeling (p. 216). Kiev: Naukova dumka. http://www.gmdh.net/articles/theory/bookNoiseIm.pdf
6. Степашко В.С., Єфіменко С.М., Савченко Є.А. Комп’ютерний експеримент в індуктивному моделюванні. – Київ: Наукова думка. – 2014. – 222 с.
7. Quenouille, M. H. (September 1949). Problems in Plane Sampling. The Annals of Mathematical Statistics 20 (3): 355–375.
8. Quenouille, M. H. (1956). Notes on Bias in Estimation. Biometrika 43 (3-4): 353–360
9. Tukey, J. W. (1958). Bias and confidence in not quite large samples. The Annals of Mathematical Statistics 29: 614–623
10. Bootstrap Methods: Another Look at the Jackknife." Ann. Statist. 7 (1) 1 - 26, January, 1979.
11. Ивахненко А. Г. Метод группового учета аргументов — конкурент метода стохастической аппроксимации. — Автоматика, 1968, No 3, с. 58—72

12. Шелудько О.И., Патереу С.Г. Упрощенный алгоритм идентификации характеристик сложных объектов по МГУА. - В кн. «Программы прямого синтеза моделей (по принципу самоорганизации). Киев, РФАП 1975, вып. 1.
13. В. С. Степашко, О. С. Булгакова, В. В. Зосімов. Ітераційні алгоритми індуктивного моделювання: [монографія]. Київ : Наукова думка, 2018.
14. D. K. Faddeev, V. N. Faddeeva, “Computational methods of linear algebra”, Computational methods of linear algebra. Parallel computing, Zap. Nauchn. Sem. LOMI, 54, "Nauka",1975, 3–228.
15. Fukunaga, K. (1990). Introduction to statistical pattern recognition (p. 491). Boston: Academic Press.
16. Shakhnarovich, G., Fisher, J., & Darrell, T. (2002). Face recognition from long-term observations. In Proceedings of 7th European Conference on Computer Vision (pp. 851-868). Copenhagen, Denmark: Springer, Berlin, Heidelberg.
17. Aranjelovic, O., Shakhnarovich, G., Fisher, J., Cippola, R., & Darrell, T. (2005). Face recognition with image sets using manifold density divergence. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 581--588). Piscataway, New Jersey: IEEE.
18. Ning, Xia & Karypis, George. (2009). The Set Classification Problem and Solution Methods. Proceedings - IEEE International Conference on Data Mining Workshops, ICDM Workshops 2008. 720 - 729. 10.1109/ICDMW.2008.113.
19. Ie. Nastenko, O. Konoval, O. Nosovets, V. Pavlov. Set Classification, pp. 44-83 - In: Techno-Social Systems for Modern Economical and Governmental Infrastructures (Advances in Finance, Accounting, and Economics). : IGI Global; 1 ed. (July 13, 2018), 351 P.
20. Voitikova, M., & Khursa, R. (2013). Linear regression modeling of blood pressure data to determine the risk of acute hypotensive episodes. News of the Academy of Sciences of Belarus, 1, 117-122.

21. Nastenko, Ie., Pavlov, V., & Nosovets, O. (2014). Synthesis classifiers differential diagnosis of the circulatory systems pathological conditions in a multidimensional space. Inductive Modeling of Complex Systems, 6, 105-122. (in Russian)
22. Biuk, Z., & Loncaric, S. (2001). Face recognition from multi-pose image sequence. In Proceedings of 2nd Int'l Symposium on Image and Signal Processing and Analysis (pp. 319-324). Pula, Croatia: IEEE.
23. Shakhnarovich, G., Fisher, J., & Darrell, T. (2002). Face recognition from long-term observations. In Proceedings of 7th European Conference on Computer Vision (pp. 851-868). Copenhagen, Denmark: Springer, Berlin, Heidelberg.
24. Samsudin, N., & Bradley, A. (2010). Nearest neighbour group-based classification. Pattern Recognition, 43(10), 3458-3467.
25. Wang W., Ozolek, J., Slepčev, D., Lee, A., Cheng Ch., & Rohde, G. (2011). An optimal transportation approach for nuclear structure-based pathology. IEEE Transactions on Medical Imaging, 30(3), 621-631. http://dx.doi.org/10.1109/tmi.2010.2089693
26. Jung, S., & Qiao, X. (2014). A statistical approach to set classification by feature selection with applications to classification of histopathology images. Biometrics, 70(3), 536-545. http://dx.doi.org/10.1111/biom.12164
27. Aranjelovic, O., Shakhnarovich, G., Fisher, J., Cippola, R., & Darrell, T. (2005). Face recognition with image sets using manifold density divergence. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 581--588). Piscataway, New Jersey: IEEE.
28. Tan, H., & Gao, Y. (2017). Patch-based principal covariance discriminative learning for image set classification. IEEE Access, 5, 15001-15012. http://dx.doi.org/10.1109/access.2017.2733718
29. Wang, R., Guo, H., Davis, L., & Dai, Q. (2012). Covariance discriminative learning: A natural and efficient approach to image set classification. In Proceedings of Computer Vision and Pattern Recognition (pp. 2496-

2503). Washington, DC, USA: IEEE.

30. Zabara, S., Filimonova, N., & Zelensky, K. (2009). The method of selection signals invariant features. Reports of the Ukraine's National Academy of Sciences, 3, 55-60.
31. Moody, G., & Lehman, L. (2009). Predicting acute hypotensive episodes: the10th annual PhysioNet/Computers in Cardiology Challenge. Computers in Cardiology, 36, 541-544.
32. Knyshov, G., Nastenko, I., Kondrashova, N., Nosovets, O., & Pavlov, V. (2014). Combinatorial algorithm for constructing a parametric feature space for the classification of multidimensional models. Cybernetics and Systems Analysis, 50(4), 627-633. http://dx.doi.org/10.1007/s10559-014-9651-3
33. Stepashko, V. (1981). A combinatorial algorithm of the group method of data handling with optimal model scanning scheme. Soviet Automatic Control, 14(3), 24-28.
34. Павлов В.А., Классификация объектов, заданных временными рядами. «Індуктивне моделювання складних систем», збірник наук. праць, №5 , – К.: МННЦІТС, 2013. – С.135-142