

УДК 681.5.015

## ДОСЛІДЖЕННЯ СЕЛЕКЦІЙНИХ ВЛАСТИВОСТЕЙ АЛГОРИТМУ MULTI

С.М.Єфіменко

*Міжнародний науково-навчальний центр інформаційних технологій та систем (МННЦ ІТС) НАН та МОН України,  
[syefim@ukr.net](mailto:syefim@ukr.net)*

У статті виконано дослідження селекційних властивостей алгоритму MULTI та представлено його удосконалену модифікацію MULTI-R. Основними особливостями модифікації є використання рекурентних процедур оцінювання параметрів та додаткова оптимізація структур моделей. Наведено результати експериментального дослідження алгоритму.

*Ключові слова:* індуктивне моделювання, МГУА, комбінаторний алгоритм, багатоетапний алгоритм, рекурентні обчислення.

The paper analyzes selective properties of the MULTI GMDH algorithm and presents its improved and revised version MULTI-R. This new version differs from the original algorithm by using a recurrent procedure of parameters estimation and additional optimizing the model structure. The numerical characteristics of the processing speed and structural accuracy of this algorithm are given

*Keywords:* inductive modelling, GMDH, combinatorial algorithm, multistage algorithm, recurrent computations.

В статье выполнено исследование селекционных свойств алгоритма MULTI и представлена его усовершенствованная модификация MULTI-R. Основными особенностями модификации является использование рекуррентных процедур оценивания параметров и дополнительная оптимизация структур моделей. Приведены результаты экспериментального исследования алгоритма.

*Ключевые слова:* индуктивное моделирование, МГУА, комбинаторный алгоритм, многоэтапный алгоритм, рекуррентные вычисления.

### 1. Вступ

Алгоритми МГУА [1-5] є ефективним засобом побудови моделей складних об'єктів і процесів за даними спостережень.

Одним з найпопулярніших алгоритмів МГУА є комбінаторний алгоритм СОМВІ [6-8], що виконує повний перебір структур моделей та знаходить найкращу за значенням заданого зовнішнього критерію. Обчислювальна складність алгоритму має експоненційну залежність від кількості аргументів, тому час моделювання збільшується удвічі при додаванні одного аргументу. Навіть застосування спеціальних засобів (наприклад, розпаралелювання обчислень) не дозволяє розв'язувати задачі із більш ніж 40 входами (аргументами) [9].

Мета алгоритмів послідовного пошуку – знайти оптимальну модель, отриману за допомогою повного перебору. Одним із таких алгоритмів є

багатоетапний комбінаторно-селекційний алгоритм MULTI [10], що ґрунтується на принципі неостаточних рішень та детально розглядається в роботі [11]. За рахунок скорочення пошуку алгоритм дозволяє розв'язувати задачі моделювання із більш ніж 100 аргументами. До його недоліків можна віднести те, що функція мінімальних значень критерію селекції може бути багатоестремальною (як це показано далі у цій роботі), що істотно впливає на правило зупинки алгоритму, а відтак, і на пошук найкращої моделі.

Мета цієї роботи – удосконалити комбінаторно-селекційний алгоритм MULTI за рахунок:

- застосування рекурентних процедур оцінювання параметрів моделей;
- додаткової оптимізації структур моделей;
- коригування правила зупинки алгоритму.

## 2. Постановка задачі

У загальному випадку задача структурно-параметричної ідентифікації полягає у формуванні за даними вибірки деякої множини моделей різної структури виду

$$\hat{y}_f = f(X, \hat{\theta}_f) \quad (1)$$

і пошуку оптимальної моделі за умовою

$$f^* = \mathbf{arg\,min}_{f \in \tilde{\mathcal{S}}} CR(y, f(X, \hat{\theta}_f)), \quad (2)$$

де оцінки параметрів в (2) є розв'язком ще однієї екстремальної задачі виду

$$\hat{\theta}_f = \mathbf{arg\,min}_{\theta_f \in R^{s_f}} QR(y, X, \theta_f, s_f), \quad (3)$$

де  $s_f$  називається складністю моделі  $f$  і дорівнює числу ненульових компонент в моделі (2);  $QR$  - критерій якості розв'язку задачі параметричної ідентифікації кожної окремої моделі, яка генерується в задачі структурної ідентифікації, де моделі порівнюються за критерієм  $CR$ . При цьому критерії  $QR$  і  $CR$  повинні бути різними, оскільки (2) є задачею дискретної, а (3) - неперервної оптимізації.

## 3. Комбінаторно-селекційний алгоритм MULTI

Алгоритм MULTI був побудований для скорочення пошуку результату повного перебору моделей. Це селекційний алгоритм, що має скінченне число етапів (не більше ніж кількість незалежних змінних  $m$ ). Головною особливістю алгоритму є використання принципу неостаточних рішень. Згідно з цим принципом декілька моделей (а не одна найкраща) відбираються і передаються на наступний етап алгоритму. Це дозволяє підвищити ймовірність знаходження глобального мінімуму критерію селекції  $CR$ .

Загальна схема алгоритму є такою [11]: на першому етапі будуються всі моделі з одним аргументом, і всі вони відбираються; на другому етапі до

відібраних моделей додаються лише відсутні в них аргументи, і вибираються найкращі моделі складності 2, які покращують значення критерію. Ця процедура виконується до тих пір, поки значення критерію не зменшиться. Основними властивостями оригінального багатоетапного алгоритму MULTI є такі [11]:

- алгоритм є комбінаторно-селекційним, що означає, що на кожному кроці а) виконується повний перебір всіх аргументів, що додаються послідовно в частинні моделі, і б) задане число найкращих моделей (серед усіх згенерованих) відбираються на наступний етап;

- алгоритм має обмежену кількість етапів, і його основні параметри, такі як правило зупинки і свобода вибору на кожному етапі, визначаються автоматично;

- кількість аргументів (складність) частинних моделей збільшується на один на кожному етапі;

- алгоритм зупиняється після знаходження істинної моделі (у випадку незашумлених даних) або моделі оптимальної складності (для зашумлених даних);

- алгоритм має поліноміальну складність, тоді як складність повного перебору – експоненційна.

Процедура послідовної генерації структур частинних моделей дозволяє використовувати рекурентні процедури оцінювання параметрів. Наприклад, послідовність всіх комбінацій структур моделей другого етапу для задачі з  $m = 5$  аргументів (та найкращої моделі  $y_1 = a_1x_1$  першого етапу) буде такою:

$$y_{21} = a_1x_1 + a_2x_2$$

$$y_{22} = a_1x_1 + a_3x_3$$

$$y_{23} = a_1x_1 + a_4x_4$$

$$y_{24} = a_1x_1 + a_5x_5$$

Очевидно, що параметри зазначених моделей другого етапу можна оцінити за рекурентною процедурою (на основі оцінок параметрів, отриманих на першому етапі алгоритму). Подібна процедура застосовується і на третьому етапі до будь-якої з найкращих моделей, відібраних на другому етапі, і т.д.

#### 4. Рекурентні обчислення

Одним з найбільш ефективних способів розширення можливостей перебірних алгоритмів (у сенсі підвищення їх швидкодії) є використання рекурентних обчислень [12-14]. Обчислювальна складність етапу оцінювання параметрів пропорційна другому степеню складності моделі (кількості оцінюваних параметрів) для будь-якого рекурентного алгоритму і третьому степеню для нерекурентного.

Обчислювальна суть модифікації алгоритму Гаусса полягає в наступному. На  $s$ -му кроці ( $s = \overline{1, m}$ ) нормальна матриця  $H_s = X_s^T X_s$  (для відомої формули оцінки параметрів за МНК) розмірності  $s \times s$  зводиться до верхнього трикутного вигляду обчисленням елементів  $h_{i,s}^s, i = 2, s-1$  і  $h_{s,i}^s, i = \overline{2, s}$ , в той час як елементи вкладеної матриці  $H_{s-1}$  розмірності  $(s-1) \times (s-1)$ , зведеної до верхнього трикутного вигляду на попередньому кроці, залишаються незмінними. Таким чином, на  $s$ -му кроці виконується обчислення "обрамлення" до елементів розширеної матриці системи нормальних рівнянь, обчислених на попередньому кроці, та відповідного елемента  $g_s = X_s^T y$  у цієї матриці:

$$\begin{bmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1,s-1} & \mathbf{h}_{1s} & \dots & h_{1m} & \vdots & g_1 \\ h_{21} & h_{22} & h_{23} & \dots & h_{2,s-1} & \mathbf{h}_{2s} & \dots & h_{2m} & \vdots & g_2 \\ h_{31} & h_{32} & h_{33} & \dots & h_{3,s-1} & \mathbf{h}_{3s} & \dots & h_{3m} & \vdots & g_3 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \vdots & \dots \\ h_{s-1,1} & h_{s-1,2} & h_{s-1,3} & \dots & h_{s-1,s-1} & \mathbf{h}_{s-1,s} & \dots & h_{s-1,m} & \vdots & g_{s-1} \\ \mathbf{h}_{s1} & \mathbf{h}_{s2} & \mathbf{h}_{s3} & \dots & \mathbf{h}_{s,s-1} & \mathbf{h}_{ss} & \dots & h_{sm} & \vdots & \mathbf{g}_s \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \vdots & \dots \\ h_{m1} & h_{m2} & h_{m3} & \dots & h_{m,s-1} & h_{ms} & \dots & h_{mm} & \vdots & g_m \end{bmatrix}$$

Особливістю запропонованої модифікації є те, що інформація, необхідна для включення наступного аргументу, зберігається в тій же матриці  $H$  нормальної системи, отже, метод не потребує використання додаткової оперативної пам'яті комп'ютера.

### 5. Удосконалений комбінаторно-селекційний алгоритм MULTI-R

Оригінальний алгоритм MULTI реалізує процедуру послідовного збільшення кількості аргументів і не передбачає виключення нерелевантних аргументів. Однак у випадку великої кількості істинних аргументів та їх складних множинних залежностей на початкових етапах алгоритму існує ймовірність включення в найкращі моделі надлишкових нерелевантних аргументів. Тому алгоритм вимагає додаткової оптимізації кращих вибраних структур моделей. Для вирішення цієї проблеми був побудований удосконалений алгоритм MULTI-R. На кожному  $s$ -му етапі алгоритму виконуються такі операції:

1. Частинні моделі генеруються шляхом додавання до кожної  $i$ -ої моделі попереднього етапу одного з відсутніх аргументів (що відповідають нульовому елементу структурного вектора  $d_i$  цієї моделі). Для забезпечення неповторюваності структур використовується спеціальна процедура перевірки, під час якої аналізуються відібрані найкращі моделі попереднього етапу.

Виконання цієї процедури починається з другого етапу, оскільки кожна згенерована структура моделі першого етапу не повторюється.

2. Параметри кожної неповторюваної моделі  $\hat{y}_s$  оцінюються за МНК.

3. Відбирається  $F$  найкращих за значенням критерію моделей (від усіх, що генеруються на етапі  $s$ ), які передаються на наступний етап. Для отримання результату повного перебору (як показано в [11]) значення  $F$  достатньо взяти рівним кількості аргументів  $m$ .

4. Після завершення всіх  $m$  етапів алгоритму (із включенням відсутніх аргументів) виконується зворотний хід. Із кращої моделі (нехай вона має складність  $l$ ,  $l \leq m$ ) на кожному кроці  $i$ ,  $i = m, 1$ , послідовно виключається один аргумент і на наступний етап ( $i-1$ ) передається модель з мінімальним значенням критерію селекції. Найкраща модель алгоритму відбирається саме під час зворотного ходу.

На рис. 1 представлена блок-схема алгоритму MULTI-R.

## 6. Результати тестового експерименту

З метою визначення ефективності розробленої модифікації проведено експериментальне дослідження властивостей алгоритму MULTI-R за допомогою тестової задачі. Вимірювався час моделювання за умови знаходження результату повного перебору за допомогою алгоритму COMBI.

Тестова задача формувалася таким чином: генерувалася матриця  $X$  розмірності  $50 \times 25$  (50 спостережень для 25 аргументів) для системи умовних рівнянь  $X\theta = y$ . Вектор  $y$  формувався як лінійна комбінація перших двадцяти аргументів з додаванням рівномірно розподіленого шуму. Результат моделювання порівнювався з отриманим за допомогою COMBI.

Найкраща модель за алгоритмом COMBI містила всі істинні аргументи та додатковий  $x_{23}$ . Значення критерію для цієї моделі:  $AR_1 = 1,167$ . Час моделювання склав 199 сек.

Найкраща модель за прямим ходом алгоритму MULTI-R містила всі аргументи, окрім  $x_{22}$ . Значення критерію:  $AR_2 = 1,225$ . Час моделювання склав 0,03 сек.

Використання зворотного ходу дозволило отримати результат повного перебору.

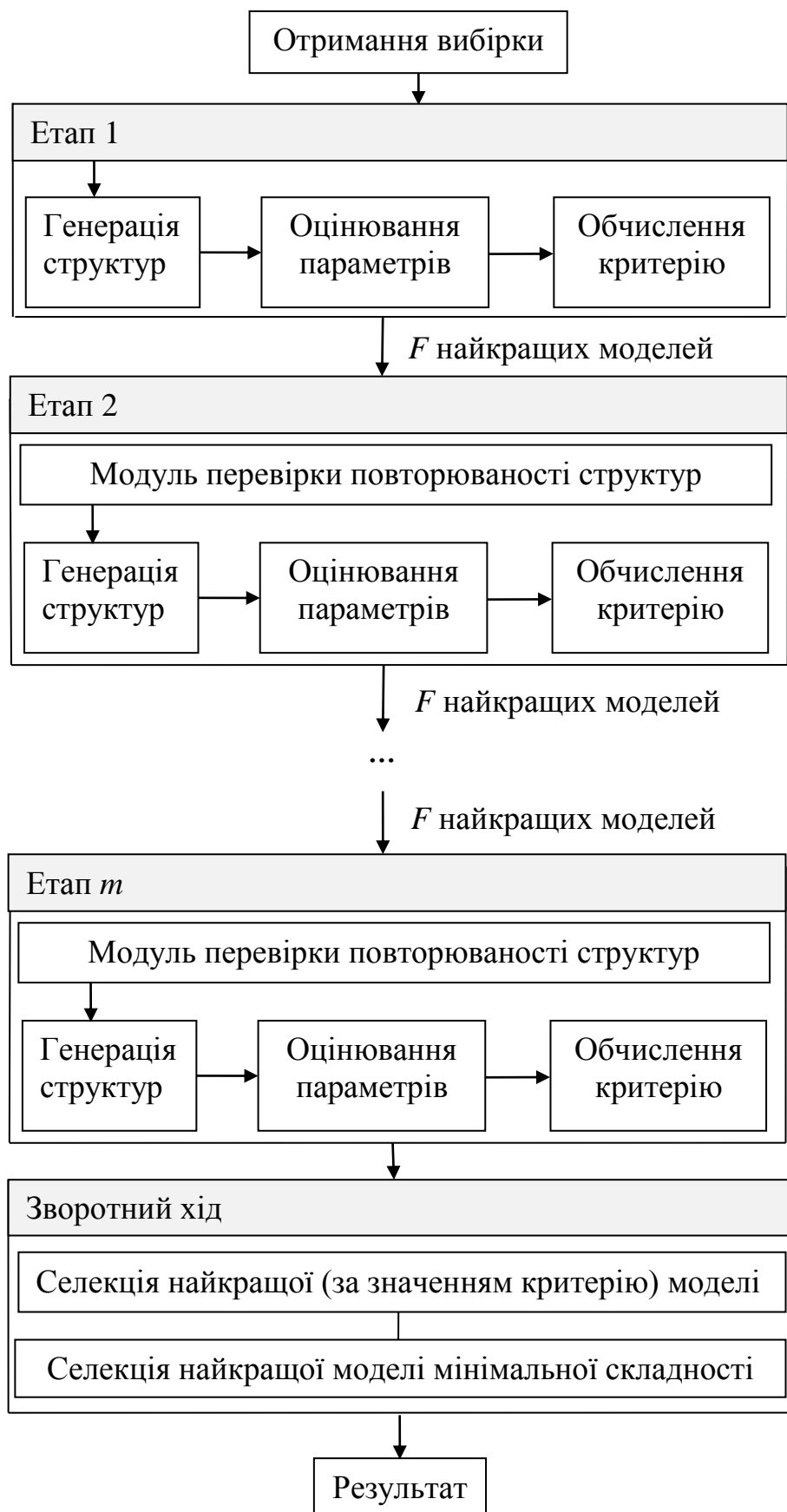


Рис. 1 Блок-схема алгоритму MULTI-R

На рисунку 2 показані мінімальні і максимальні значення критерію регулярності для всіх побудованих моделей при послідовному виконанні етапів.

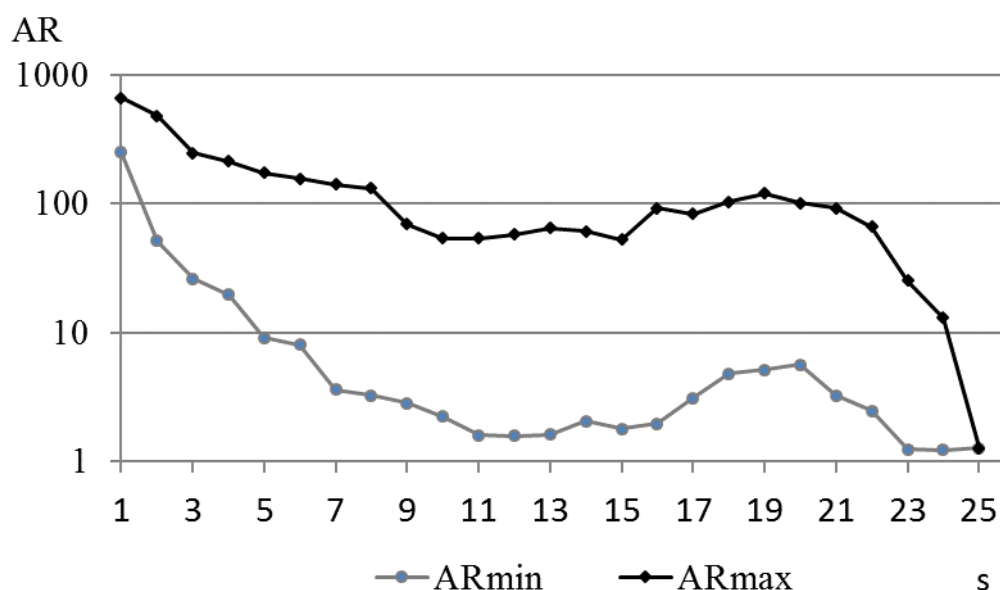


Рис. 2 Максимальні та мінімальні значення критерію регулярності (в логарифмічній шкалі)

Представлена в цій роботі модифікація відрізняється від оригінального алгоритму MULTI відсутністю правила зупинки. Вона має перевагу тоді, коли крива мінімальних значень критерію селекції є багатоекстремальною, що продемонстровано на рис. 2. Оригінальний алгоритм MULTI при розв'язанні цієї тестової задачі зупиниться на етапі 12 і вибере модель, що містить щонайбільше 12 з 20 істинних аргументів.

## Висновки

Виконано дослідження селекційних властивостей алгоритму MULTI та розроблено удосконалений комбінаторно-селекційний алгоритм послідовного пошуку моделі оптимальної складності MULTI-R. Основними особливостями модифікації є використання рекурентних процедур оцінювання параметрів та додаткова оптимізація структур моделей.

Результати тестового експерименту показали, що процедура оптимізації структур моделей в алгоритмі MULTI-R дозволяє отримати результат повного перебору, отриманого за алгоритмом COMBI.

У якості недоліку алгоритму можна вказати збільшення часу моделювання, яке компенсується використанням рекурентних обчислень.

## Література

1. A.G. Ivakhnenko, "Heuristic Self-Organization in Problems of Automatic Control," *Automatica (IFAC)*, 1970, no. 6, pp. 207-219.
2. Self-organizing methods in modeling: GMDH type algorithms / Ed. S.J. Farlow, New York, Basel: Marcel Decker Inc., 1984, 350 p.
3. H.R. Madala, A.G. Ivakhnenko, *Inductive Learning Algorithms for Complex Systems Modeling*, New York: CRC Press, 1994, 368p.
4. M. Snorek, P. Kordik, "Inductive Modelling World Wide the State of the Art," *Proc. of 2nd Int. Workshop on Inductive Modelling*, Prague: CTU, 2007, pp. 302-304, Sept. 2007
5. V. Stepashko, "Developments and Prospects of GMDH-Based Inductive Modeling," In: *Advances in Intelligent Systems and Computing II* / N. Shakhovska, V. Stepashko, eds, AISC book series, vol. 689, Cham: Springer, 2018, pp. 474-491.
6. V.S. Stepashko, "A Combinatorial Algorithm of the Group Method of Data Handling with Optimal Model Scanning Scheme", *Soviet Automatic Control*, 1981, vol.14, no. 3, pp. 24-28.
7. A.G. Ivakhnenko, J.-A. Müller, "Recent Developments of Self-Organising Modeling in Prediction and Analysis of Stock Market," *Microelectronics Reliability*, 1997, no. 37, pp. 1053–1072.
8. L. Anastasakis, N. Mort, "The development of self-organization techniques in modelling: a review of the group method of data handling (GMDH)," *Research Report 813, Department of Automatic Control & Systems Engineering, The University of Sheffield, UK*, 2001, 38 p.
9. S. Yefimenko, V. Stepashko, "Intelligent Recurrent-and-Parallel Computing for Solving Inductive Modeling Problems," *Proceedings of 16th International Conference on Computational Problems of Electrical Engineering (CPEE-2015)*, September 2-5, 2015, Lviv, Ukraine, 2015, pp. 236-238.
10. V.S. Stepashko, "A Finite Selection Procedure for Pruning an Exhaustive Search of Models," *Soviet Automatic Control*, 1983, vol. 16, no. 4, pp. 84-88.
11. Степашко В.С., Костенко Ю.В. Комбинаторно-селекционный алгоритм последовательного поиска модели оптимальной сложности // *Праці І Міжн. конф. з індуктивного моделювання*, Львів, 20–25 травня, 2002. – Т. 1., ч. 1. – Львів: ДНДІІ, 2002. – С. 72–76.
12. G.A.F. Seber, *Linear Regression Analysis*, New York – London – Sydney – Toronto: John Wiley and Sons, 1977, 496 p.
13. J. Gergely, "Matrix inversion and the solution of systems of linear and non-linear equations by the method of bordering", *USSR Computational Mathematics and Mathematical Physics*, vol. 19, iss. 4, 1979, pp. 1-10.
14. V.S. Stepashko, S.N. Efimenko, "Sequential Estimation of the Parameters of Regression Model," *Cybernetics and Systems Analysis*, New York: Springer, July, 2005, vol. 41, no. 4, pp.631-634.