УДК 6:004.8

# ОЦІНКА ОБЧИСЛЮВАЛЬНОЇ СКЛАДНОСТІ КОМБІНАТОРНО-ГЕНЕТИЧНОГО АЛГОРИТМУ КОМБІ-ГА

О.Г. Мороз

*Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, Київ, пр-т Академіка Глушкова, 40*

*olga_moroz@irtc.org.ua*

В статті представлено розрахунок обчислювальної складності алгоритму КОМБІ-ГА та порівняння його ефективності з алгоритмами КОМБІ, MULTI та LASSO.
*Ключові слова: Комбінаторний алгоритм МГУА, перебірний алгоритм КОМБІ-ГА, складність алгоритму, алгоритм MULTI, алгоритм LASSO.*

The article presents the calculation of the computational complexity of the COMBI-GA algorithm and comparison its effectiveness with algorithms COMBI, MULTI and LASSO.
*Keywords: Combinatorial GMDH algorithm, sorting-out algorithm COMBI-GA, algorithm complexity, algorithm MULTI, algorithm LASSO.*

В статье представлен расчет вычислительной сложности алгоритма КОМБИ-ГА и сравнение его эффективности с алгоритмами КОМБИ, MULTI и LASSO.
*Ключевые слова: Комбинаторный алгоритм МГУА, переборный алгоритм КОМБИ-ГА, сложность алгоритма, алгоритм MULTI, алгоритм LASSO.*

## Вступ

Перебірний комбінаторно-генетичний алгоритм КОМБІ-ГА [1, 2] є ефективним засобом розв’язання задач побудови моделей об'єктів і процесів за даними спостережень в умовах невизначеності та низького рівня апріорних знань про модельований об’єкт.

В його основі лежить комбінаторний алгоритм МГУА (КОМБІ) [3 – 5] з повним перебором варіантів структур моделей для пошуку найкращої. Цей алгоритм обмежений у практичному застосуванні, оскільки потребує великих витрат обчислюваних ресурсів і часу та без використання спеціальних засобів не дозволяє розв’язувати задачі з більш ніж 30-ма вхідними змінними (аргументами) навіть на сучасних комп’ютерах. Кількість перебраних моделей за цим алгоритмом є показниковою функцією від кількості незалежних змінних, тобто зростає експоненційно.

Альтернативним найбільш ефективним перебірним алгоритмом МГУА є MULTI [6 – 8], який має поліноміальну складність.

Одним з найперспективніших регресійних методів для пошуку оптимальної моделі є сучасний алгоритм LASSO [9].

В цій роботі оцінюється теоретична складність алгоритму КОМБІ-ГА та на тестових прикладах різної розмірності порівнюється ефективність цього алгоритму та алгоритмів КОМБІ, MULTI і LASSO.

## Гібридний перебірний алгоритм КОМБІ-ГА

Алгоритм КОМБІ-ГА поетапно формує множину найбільш перспективних структур частинних моделей і знаходить оптимальну з них, використовуючи генетичні оператори селекції, кросинговеру та мутації, які визначають механізм перебору. Формально цей алгоритм можна описати так:

$$\text{КОМБІ-ГА} = \{\boldsymbol{Z}, \boldsymbol{y}, \boldsymbol{f}, \boldsymbol{X}, D, CR, \boldsymbol{P}_0, H, M, \boldsymbol{G}, k, F\},$$

де $\boldsymbol{Z}$[n×r] – матриця початкових вимірювань, n – кількість спостережень, r – кількість початкових вхідних змінних;
$\boldsymbol{y}$[n×1] – вектор вимірювань вихідної змінної;
$\boldsymbol{f}$[m×1] – вектор заданих $m$ базисних функцій вхідних змінних, $m \geq r$;
$\boldsymbol{X}$[n×m] – матриця вимірювань базисної множини аргументів, $m \geq r$;
$D$ – правило поділу $\boldsymbol{X}$[n×m] і $\boldsymbol{y}$[n×1] на тестову $A$ та перевірну $B$ частини;
$CR$ – зовнішній критерій (або цільова функція);
$\boldsymbol{P_0}$ – множин бінарних хромосом (закодованих структур моделей) початкової популяції ГА;
$H$ – розмір початкової популяції, $H < 2^m$-1;
$M$ – розмір поточної популяції, $M \leq H$;
$\boldsymbol{G}$ – множина генетичних операторів;
$k$ – правило зупинки ГА;
$F$ – кількість кращих моделей (свобода вибору) на кожній ітерації, $F<M$.

Загальну Блок-схему алгоритму подано на рис. 1.

Детально цей алгоритм описано в [2], де подаються й обговорюються результати дослідження його ефективності на тестових і реальних задачах.

## Оцінювання обчислювальної складності алгоритму КОМБІ-ГА

Під обчислювальною складністю алгоритму зазвичай розуміють деяку функцію залежності обсягу виконаної ним роботи від розміру вхідних даних, яку можна отримати підрахунком кількості виконуваних арифметичних операцій. Її оцінка дозволяє передбачити час роботи конкретного алгоритму і порівнювати ефективність різних алгоритмів розв’язання однакових задач.

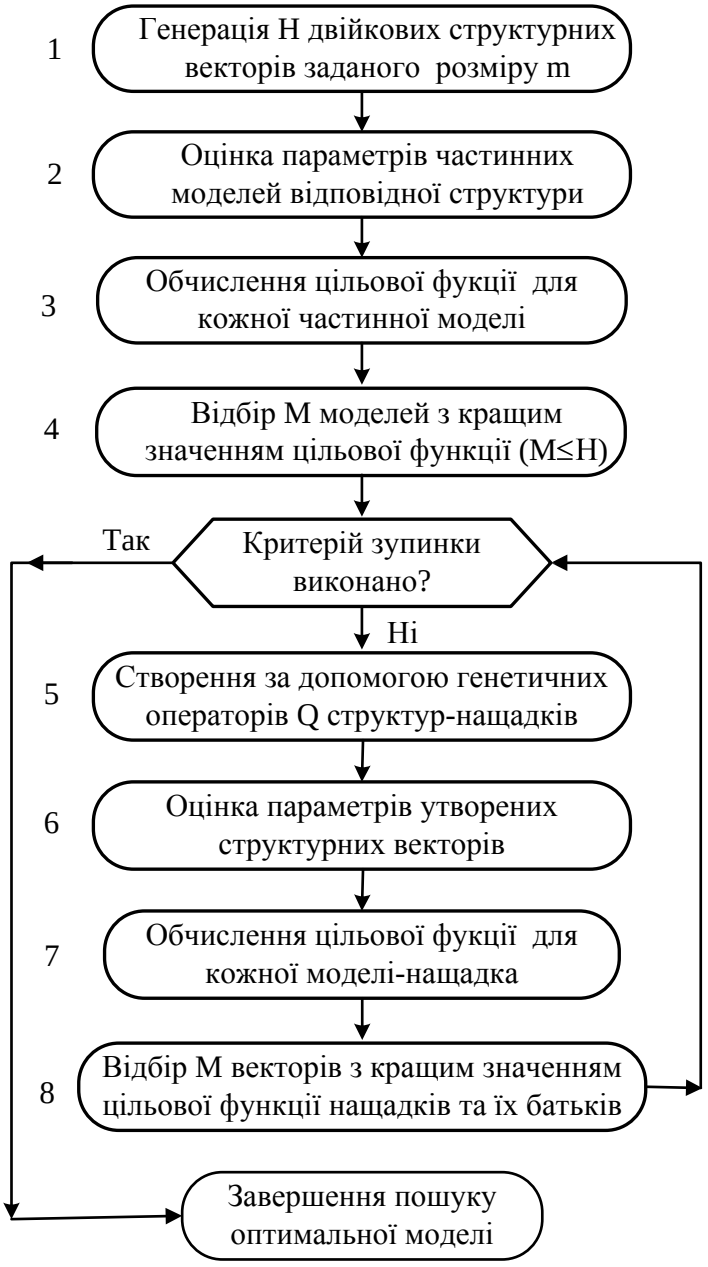


Рис. 1 Генетичний пошук оптимальної моделі

В [10, 11] представлено обчислювальна складність найбільш ефективних перебірних алгоритмів МГУА, таких як MULTI, BSS та ін.

Підрахуємо обчислювальну складність алгоритму КОМБІ-ГА з урахуванням найбільш ресурсно-затратних його елементів: оцінювання вектора параметрів $\boldsymbol{\theta}$ і розрахунок значення критерію регулярності кожної частинної моделі.

Розроблений у цій роботі новий комбінаторно-генетичний алгоритм КОМБІ-ГА за одну ітерацію перебирає (тобто обчислює ЦФ) максимум $2C_M^2 = M(M-1)$ структурних векторів-нащадків, отриманих за допомогою

генетичних операторів кросинговеру та мутації з структурних векторів-батьків поточної популяції розміром *M*.

Після обчислення ЦФ усіх $M(M-1)$ структурних векторів-нащадків із загальної кількості структурних векторів поточної ітерації, що рівне $M + M(M-1) = M^2$, оператором селекції створюється нова поточна популяція розміром *M* (ЦФ структурних векторів якої – відомо) і починається нова ітерація. Отже, для кількості моделей *QM*, порівнюваних алгоритмом КОМБІ-ГА за 1 ітерацію, виконується нерівність:

$$QM \le 2C_M^2 = M(M-1), \tag{1}$$

Тоді за 1 запуск алгоритму КОМБІ-ГА кількість переглянутих моделей $QM_T$ залежно від кількості ітерацій *T* генетичного алгоритму та розміру початкової популяції *H* обмежена нерівністю:

$$QM_T \le H + T \cdot M(M-1) \le T \cdot M(M-1). \tag{2}$$

На кожній ітерації необхідно оцінити вектор параметрів $\hat{\theta}_A$ для кожного структурного вектора максимальної розмірності $m \times 1$ на підвибірці *A* і обчислити вектор оцінки вихідної величини $\hat{y}_B$ на частині *B*:

$$\hat{y}_i = \sum_{j=1}^{m} \hat{\theta}_{Aj}\, x_{ij},\ i = \overline{1, n_B}\ . \tag{3}$$

Після цього слід обчислити значення критерію регулярності:

$$AR = \frac{1}{n_B} \sum_{j=1}^{n_B} (y_{Bj} - \hat{y}_{Bj})^2. \tag{4}$$

Кількість операцій для реалізації формули (3) за заданих оцінок $\hat{\theta}_{Aj}$ та $\hat{y}_{Bi}$ рівна $2n_B(m-1)$, а для формули (4) $2n_B+1$. Отже, для обчислення критерію регулярності потрібно виконати таку кількість операцій (без урахування затрат на оцінювання параметрів):

$$2n_B(m-1) + 2n_B + 1 = 2mn_B + 1\ .$$

Для отримання значень $\hat{\theta}_j$ методом Гауса необхідно $m^2 + (2/3)\, m^3$ операцій [11].

Отже, оцінка числа операцій *OQ*, необхідних для побудови однієї моделі з *m* аргументами та оцінки її якості, яка включає в себе кількість

операцій для оцінювання параметрів за МНК, тобто для розв'язання нормальної системи лінійних алгебраїчних рівнянь розміром $m \times m$ та обчислення критерію регулярності, є такою:

$$OQ \le 2mn_B + m^2 + 2/3m^3 \le 2mn_B + m^2 + m^3 = m[2n_B + m(m+1)]. \quad (5)$$

З (2) та (5) одержимо таку оцінку складності $S$ алгоритму:

$$S \le T \cdot M(M-1)(2mn_B + m^2 + m^3). \quad (6)$$

Отже, складність $S$ має порядок $O(m^3)$.

З аналізу методу оцінювання коефіцієнтів і розрахунку критерію селекції очевидно, що обчислювальна складність алгоритму КОМБІ-ГА має поліноміальну складність $m^3$ за кількістю аргументів, як і в кращих перебірних алгоритмах МГУА, зокрема в MULTI. Отже, можна стверджувати, що КОМБІ-ГА теоретично є не менш ефективним, ніж інші відомі перебірні алгоритми індуктивного моделювання.

## Порівняння ефективності алгоритмів на тестових експериментах

Тестові експерименти проводилися для порівняння обчислювальної ефективності алгоритмів КОМБІ, MULTI, LASSO та КОМБІ-ГА. При цьому, зокрема, для КОМБІ-ГА застосовувалась така методика тестування:

1) генерація матриці вхідних аргументів $\boldsymbol{X}[n \times m]$, обчислення вектора вихідної змінної $\boldsymbol{y}[n \times 1]$ та задання рівня шуму;
2) поділ вибірки $\boldsymbol{W}(\boldsymbol{X}, \boldsymbol{y})$ на навчальну, перевірну та екзаменаційну підвибірки;
3) встановлення кількості запусків алгоритму КОМБІ-ГА;
4) генерація початкової популяції ГА заданого розміру M;
5) задання значень параметрів ГА і генетичних операторів;
6) знаходження кращих моделей;
7) оцінка якості отриманих результатів та їх візуалізація.

При цьому було встановлено такі характеристики проведених тестових експериментів:

- вибірка вхідних даних ділиться на навчальну, перевірну та екзаменаційну підвибірки у співвідношенні 0,5: 0,3: 0,2;
- істинні та зайві аргументи для початкової вибірки даних генеруються випадковим чином в пропорції 50:50, $x_i \in (0, 20]$; $\boldsymbol{\theta}_i \in [-6, 6]$ ;
- тестова "справжня" лінійна модель з випадковими параметрами розраховується лише з істинними аргументами;

- розмір початкової популяції ГА становить 100 хромосом (структур моделей), кількість відібраних хромосом оператором селекції 60 %, пар для схрещування 40 %;
- імовірність оператора кросинговеру $p_c = 0.8$, мутації $p_м = 0.2$;
- кількість відібраних кращих моделей $F$=5;
- кількість запусків алгоритму: 100.

В табл. 1 подано результати чисельних досліджень порівняння ефективності алгоритмів КОМБІ, MULTI, LASSO, КОМБІ-ГА. Як видно з цієї таблиці, алгоритм КОМБІ-ГА є найефективнішим. Зокрема, для тестової задачі з 1000 вхідних змінних-аргументів КОМБІ-ГА знайшов оптимальну модель у 178 разів швидше, ніж MULTI, та в 1,6 разів, ніж LASSO.

Таблиця 1

Порівняння ефективності алгоритмів

| Кількість вхідних аргументів | КОМБІ | MULTI | LASSO | КОМБІ-ГА |
|---|---|---|---|---|
| 20 | 85,2 с. | 0,48 с. | 0,53 с. | 0,01 с. |
| 200 | – | 112 с. | 3,97 с. | 2,67 с. |
| 1000 | – | 623 хв. | 5,55 хв. | 3,5 хв. |

Хоча обчислювальна складність алгоритмів MULTI і КОМБІ-ГА має однаковий порядок $O(m^3)$, час для знаходження оптимальної моделі при збільшенні кількості вхідних аргументів алгоритмом КОМБІ-ГА значно менший, що свідчить про те, що генетичний генератор структур частинних моделей є ефективнішим.

## Висновок

На сьогодні розроблено значну кількість алгоритмів для розв’язання задачі індуктивного моделювання, зокрема перебірні алгоритми МГУА з повним та неповним перебором моделей та регресійні алгоритми. Проте більшість з них потребує великих обчислювальних та часових ресурсів для пошуку оптимальної моделі.

Результати теоретичного підрахунку оцінки обчислювальної складності комбінаторно-генетичного алгоритму КОМБІ-ГА показали, що вона є поліноміальною порядку $O(m^3)$, як і в алгоритмі MULTI.

Проте за результатами чисельного порівняння часових витрат для знаходження оптимальної моделі алгоритмами КОМБІ-ГА, КОМБІ, MULTI та LASSO встановлено, що алгоритм КОМБІ-ГА працював найефективніше.

**Література**


1. Мороз О.Г. Переборный алгоритм МГУА с генетическим поиском оптимальной модели // Управляющие системы и машины. – 2016. – № 6. – С. 73-79.
2. Moroz O., Stepashko V. Hybrid Sorting-Out Algorithm COMBI-GA with Evolutionary Growth of Model Complexity / N. Shakhovska, V. Stepashko (Editors) // Advances in Intelligent Systems and Computing II, AISC book series. – Vol. 689. – Cham: Springer, 2018. – P. 346-360.
3. Степашко В.С. Оптимизация и обобщение схем перебора моделей в алгоритмах МГУА. – Автоматика. – 1979. – № 4. – С. 28-47.
4. Степашко В.С. Комбинаторный алгоритм МГУА с оптимальной схемой перебора моделей // Автоматика. – 1981. – № 3. – С. 31-36.
5. Степашко В.С. Потенциальная помехоустойчивость моделирования по комбинаторному алгоритму МГУА без использования информации о помехах. – Автоматика, 1983. – № 3. – С. 18-27.
6. Степашко В.С. Конечная селекционная процедура сокращения полного перебора моделей // Автоматика. – 1983. – № 4. – С. 84-88.
7. Степашко В.С. Костенко Ю.В. Исследование свойств комбинаторно-селекционного (многоэтапного) алгоритма МГУА // Моделирование и управление состоянием эколого-экономических систем региона. – Сб. науч. тр. – Киев: МНУЦ ИТС НАНУ, 2001. – С. 96–100.
8. Степашко В.С., Костенко Ю.В. Комбинаторно-селекционный алгоритм последовательного поиска модели оптимальной сложности // Праці І Міжнар. конф. з індуктивного моделювання, Львів, 20–25 травня, 2002. – Т. 1., ч. 1. – Львів: ДНДІІІ, 2002. – С. 72–76.
9. Tibshirani R. Regression Shrinkage and Selection via the Lasso // Journal of the Royal Statistical Society. Series B (Methodological). 1996.58, № 1, P. 267–288.
10. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. Киев: Наук. думка, 1985. – 216 с.
11. Степашко В.С., Єфіменко С.М., Савченко Є.А. Комп’ютерний експеримент в індуктивному моделюванні. – Київ: Наукова думка. – 2014. – 222 с.