

DOI <https://doi.org/10.15407/usim.2018.04.0021>
УДК 6:004.8

О.Г. МОРОЗ, мол. наук. співробітник,
Міжнародний науково-навчальний центр інформаційних технологій та систем
НАН та МОН України, просп. Глушкова, 40, Київ 03187, Україна,
olhahryhmoroz@gmail.com

ПРОГРАМНИЙ КОМПЛЕКС ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧ ІНДУКТИВНОГО МОДЕЛЮВАННЯ НА ОСНОВІ КОМБІНАТОРНО-ГЕНЕТИЧНОГО МЕТОДУ

Розроблено комп'ютерний комплекс індуктивного моделювання на основі гібридного алгоритму КОМБІ-ГА для моделювання складних систем з можливістю розв'язання дослідницьких та прикладних задач.

Ключові слова: МГУА, алгоритм КОМБІ, генетичний алгоритм ГА, гібридний алгоритм КОМБІ-ГА, програмний комплекс.

Вступ

Комбінаторно-генетичний метод КОМБІ-ГА [1] є ефективним самоорганізовним засобом індуктивного моделювання складних лінійних та нелінійних об'єктів, систем, процесів і явищ різної природи. На основі цього методу розроблено програмний комплекс засобами мови програмування *MATLAB* [2], призначений як для розв'язання практичних задач моделювання за даними спостережень в умовах неповноти інформації про об'єкт, так і для дослідження можливостей КОМБІ-ГА, зокрема, в задачах великої розмірності.

При побудові комплексу ставилися такі основні вимоги:

- наявність архітектури, яка передбачає розширення його функціональних можливостей;
- зручний інтерфейс для роботи користувача з тестовими або реальними даними;
- можливість побудови моделі різної складності та структури із застосуванням навчальної та перевіреної вибірок;
- здійснення графічного та змістовного аналізу кращих побудованих моделей;

- збереження кращих моделей в базі даних разом з проміжними розрахунками та результатами експериментів.

Далі розглянуто структуру, інтерфейс та функціональні можливості цього програмного комплексу.

Задача індуктивного моделювання

Метою задачі індуктивного моделювання складних об'єктів (систем, процесів тощо) є виявлення неявних причинно-наслідкових зв'язків і закономірностей, прихованих у даних, і подання їх в явній формі математичних моделей [3]. При цьому в багатьох випадках ця задача розв'язується за умов істотної неповноти та невизначеності даних.

Для постановки задачі індуктивного моделювання вводяться такі позначення:

$Z = [z_{ij}, i = 1, \dots, n; j = 1, \dots, r]$ — задана матриця початкових даних, де r — кількість вхідних змінних, n — кількість спостережень;

$y = (y_1, \dots, y_n)^T$ — заданий вектор вихідної змінної;

$X = [x_{ij}, i = 1, \dots, n; j = 1, \dots, m]$ — матриця перетворених початкових даних, де m — кількість еквівалентних вхідних аргументів.

Наприклад, якщо $p_i = (p_{i1}, p_{i2}, \dots, p_{ir})$ — вектор степенів усіх r змінних i -го члена повного полінома, де $p_{ij} \in \{0, 1, \dots, p\}$, і $\sum_{j=1}^r p_{ij} \leq p$ (p — максимальний степінь повного полінома), то такий поліном можна записати так:

$$y = f(X, \theta_f) = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_m x_m,$$

де θ_f — вектор невідомих параметрів, розміру $s_f \times 1$, s_f — складність моделі f (число оцінюваних параметрів). При цьому всі узагальнені «лінійні» аргументи x_i утворюють базисний набір функцій, які є нелінійними функціями початкових даних:

$$x_i = x_i(z_1, \dots, z_r) = \prod_{j=1}^{p_{ij}} z_j^{p_{ij}} \quad (i = 1, 2, \dots, m).$$

Тоді задача індуктивного моделювання складається з таких етапів [3]:

1. Оцінювання параметрів θ_f для кожної функції $f(X, \theta_f) \in \Phi$, що є розв'язком задачі неперервної оптимізації:

$$\hat{\theta} = \arg \min_{\theta_f \in R^s} QR(y, X, \theta_f),$$

де $QR(\cdot)$ — критерій параметричної ідентифікації кожної окремої моделі; $\Phi = \{f_k(X, \theta_{f_k}), k = 1, k\}$ — множина структур моделей в заданому класі.

2. Пошук на множині Φ оптимальної моделі як розв'язку задачі дискретної оптимізації за умови мінімуму зовнішнього критерію селекції $CR(\cdot)$:

$$f^* = \arg \min_{f \in \Phi} CR(y, f(X, \hat{\theta}_f)).$$

При використанні комбінаторного алгоритму КОМБІ [4] виконується перебір усіх можливих моделей, лінійних за параметрами, з вибором найкращої з них за критерієм селекції. За невеликої кількості аргументів можна виконати повний перебір, при цьому загальна кількість P_m усіх можливих моделей, які містять від одиниці до m аргументів, дорівнює $P_m = 2^m - 1$. Це показникова функція, і повний перебір моделей на сучасних однопроцесорних комп'ютерах є практично можливим, коли кількість аргументів m не більше 30, що обмежує можливості прикладного застосування КОМБІ.

Гібридний 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 самоорганізований механізм їх перебору.

Формально цей алгоритм можна описати так [1]:

$$\text{КОМБІ-ГА} = \{Z, y, f, X, D, CR, P_0, H, M, G, k, F\},$$

де $Z [n \times r]$ — матриця вимірювань вектора вхідних аргументів модельованого об'єкта, r — кількість вхідних аргументів, n — кількість точок вимірювань;

$y [n \times 1]$ — вектор вимірювань вихідної змінної модельованого об'єкту;

$f [m \times 1]$ — вектор m базисних (опорних) функцій від вхідних аргументів;

$X [n \times m]$ — матриця вимірювань базисного набору аргументів;

D — задане правило розбиття матриці вимірювань $X [n \times m]$ опорного набору аргументів (елементів вектора базисних функцій) і вектора $y [n \times 1]$ на навчальну, перевірну та екзаменаційну підвибірки.

CR — зовнішній критерій селекції (цільова функція або функція придатності ГА), на основі позначеного вище розбиття вибірки (X, y) ;

P_0 — початкова популяція ГА структур моделей у вигляді двійкових хромосом (закодованих структур частинних моделей);

H — розмір початкової популяції структур моделей, $H \ll R_m$;

M — розмір поточної популяції;

G — множина генетичних операторів;

k — критерій зупинки ГА;

F — кількість кращих частинних моделей (свобода вибору), $1 \leq F \leq M$.

Робота алгоритму КОМБІ-ГА складається з таких основних кроків:

1. Перетворення вхідних даних згідно з обраною системою базисних функцій.

2. Створення вхідної популяції КОМБІ-ГА заданого розміру M як випадкового набору частинних моделей.

3. Обчислення коефіцієнтів кожної частинної моделі з використанням методу найменших квадратів (МНК).

4. Розрахунок значень зовнішнього критерію (функції придатності ГА) для кожної моделі, наприклад, критерію регулярності, типового для МГУА.

5. Поточний відбір кращих частинних моделей. Формування нової популяції того ж розміру M .

6. Перевірка критерію зупинки — досягнення заданої точності або кількості ітерацій. Завершення алгоритму, якщо критерій виконано, інакше — перехід до наступного кроку.

7. Використання генетичних операторів (кросинговеру і мутації) з заданою ймовірністю для відібраних особин з популяції. Перехід до кроку 3.

Згідно з [5], цей алгоритм належить до класу перебірних алгоритмів МГУА з елементами випадкового пошуку.

Архітектура програмного комплексу

Загальну структуру програмного комплексу на основі алгоритму КОМБІ-ГА для розв'язання задач індуктивного моделювання показано на рис. 1.

Блок зберігання даних. Цей блок реалізовано у вигляді трьох баз даних: проміжних розрахунків, які характеризують роботу алгоритму; початкових даних, які можуть бути сформовані у два способи: генерацією тестових даних або імпорту файлу Блокнот, або *Excel*; результатів моделювання, де зберігаються результати роботи алгоритму, які можна завантажити на ПК або роздрукувати. При цьому згенеровані початкові дані також зберігаються у вигляді файлу *Excel* або Блокнот.

Завантаження початкових даних. Програмний комплекс дозволяє працювати зі заздалегідь підготовленим користувачем файлом з таблицею даних у форматах *txt* (роздільник між даними — пробіл) та *xlsx*. При цьому кожний стовпець таблиці, крім останнього, — це вектор спостережень для кожного аргумента, а в



Рис. 1. Загальна структурна схема програмного комплексу

останньому правому стовпці задаються значення вихідної величини. Дані в *Excel* мають задаватись у вигляді табл. 1.

Дані в Блокнот мають задаватись у вигляді, представленому на рис. 2.

Блок генерації даних взаємодіє з **блоком формування вибірки**, в якому згенеровані дані формують вхідну вибірку, та блоком зберігання даних, що надає користувачу можливість зберегти згенеровані дані у вигляді файлу в **Базі початкових даних**.

Таблиця 1. Формат представлення даних в таблицьному редакторі *Excel*

2,93,93	6,77	2,23	4,71	...	6,43	5,81
7,03	0,86	0,65	9,93	...	2,12	8,13
4,88	2,38	6,64	4,17	...	2,74	4,05
...
3,41	7,15	5,51	7,52	...	2,21	3,34
5,31	5,13	2,17	2,42	...	8,39	4,31

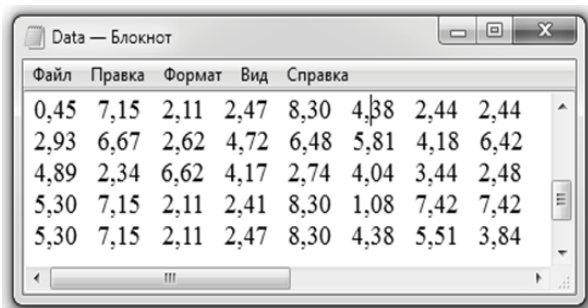


Рис. 2. Представлення даних у текстовому редакторі Блокнот

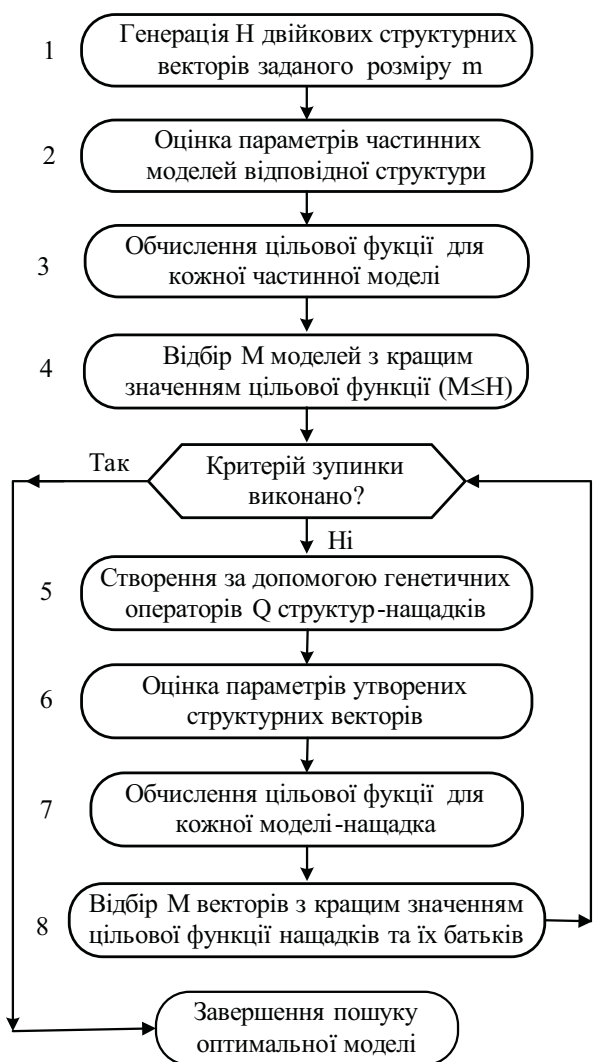


Рис.3. Генетичний пошук оптимальної моделі

Блок формування вибірки. В цьому блоці з файлу початкових даних, отриманих від «Блоку генерації даних» або «Блоку зберігання даних» у форматі *xlsx* або *txt*, відбувається зчитування даних у початкову матрицю Z та вихідний вектор y із задалегідь створеного користувачем файлу.

Блок формування задачі. У цьому блоці спершу обирається клас поліномів, у якому здійснюватиметься пошук оптимальної моделі. Далі відбувається перетворення початкових даних на вхідні, з якими працюватиме алгоритм. Тобто формується вхідна розширена матриця X та вихідний вектор y . Після цього задаються характеристики КОМБІ та генетичного генератора структур моделей, в якому генерується початкова популяція структур моделей і з якої генетичними операторами генеруються інші структури моделей. Для кожної структури моделі обчислюються значення параметрів на навчальній підвибірці. Далі на перевірній підвибірці для кожної моделі обчислюється значення критерію якості моделей.

Блок розв’язання задачі моделювання. У цьому блоці на основі даних попередніх блоків здійснюється пошук заданої кількості кращих моделей. Загальна блок-схема генетичного пошуку оптимальної моделі представлена на рис. 3. Надамо коротке пояснення до неї.

Двійкові структурні вектори генеруються в Блоці 1 рис. 3, вони є закодованими структурами частинних моделей.

Цільовою функцією, що обчислюється у Блоці 3, є, наприклад, критерій регулярності.

У Блоці 4 формується початкова батьківська популяція структурних векторів розмірності M , частина з яких утворить певну кількість випадкових пар-кандидатів для створення нащадків.

У Блоці 5 кожна з пар-кандидатів під впливом операторів кросинговеру та мутації з певною імовірністю дає двох нащадків.

Загальна кількість структур-нащадків Q — випадкове парне число, не більше $2C_M^2$. У Блоці 8 із батьків та їх нащадків формується нова (поточна) батьківська популяція структурних векторів розміру M .

Всі результати записуються в базу даних.

Інтерфейс користувача GUI в MATLAB

Одним з найважливіших чинників, які впливають на ефективність використання програмного забезпечення, є зручність інтерфейсу користувача, який є сукупністю програмних засобів, що забезпечують взаємодію користувача з комп'ютером і відповідають за введення даних і своєчасне реагування на команди користувача.

Відомі різні способи організації інтерфейсу, зокрема, із застосуванням автоматичних, діалогових та інших засобів підтримки дій користувача.

Автоматичний інтерфейс дає змогу запустити завдання на виконання, зв'язати з ним конкретні дані та виконати деякі процедури обслуговування;

Діалоговий інтерфейс — це регламентований обмін інформацією між користувачем і комп'ютером, що здійснюється в реальному масштабі часу і спрямований на спільне вирішення конкретного завдання. Кожен діалог складається з окремих процесів введення-виведення, які фізично забезпечують зв'язок між користувачем і комп'ютером.

Наразі одним з важливих атрибутів прикладної програми є графічний інтерфейс користувача (*Graphics User Interface — GUI*). Розробка графічного інтерфейсу є досить трудомісткою.

Однак для спрощення процедури створення *GUI* в *MATLAB* існують спеціальні інструментальні засоби для візуально-орієнтованого програмування і проектування додатків з *GUI*, найважливішим з яких є конструктор графічного інтерфейсу *GUIDE (GUI Designer)*.

При роботі з інструментом *GUIDE* можна створювати вікна *GUI* шляхом вибору мишею потрібних елементів управління і переміщення їх у вікно *GUI*.

Так можна створювати практично всі типові елементи інтерфейсу — кнопки, випадаючі списки, лінійки прокрутки тощо. При цьому з кожним елементом інтерфейсу пов'язаний свій фрагмент програми обробки подій, яка генерується автоматично і може коригуватися користувачем.

Інтерфейс користувача програмного комплексу в середовищі MATLAB

В розробленому програмному комплексі процес моделювання може відбуватися у двох режимах: автоматичному або діалоговому (рис. 4). В автоматичному режимі користувач завантажує в програмний комплекс лише початкові дані, після чого алгоритм автоматично запускається на виконання, використовуючи всі параметри КОМБІ-ГА, задані за замовчуванням.

У діалоговому режимі користувач може самостійно налаштовувати параметри у вікнах інтерфейсу, використовуючи за потреби Довідник, на сторінки якого можна потрапити після натискання на кнопку «Допомога».

Довідник містить всю інформацію, необхідну для заповнення всіх полів, які містить вікно, а також інструкцію з коректності введення параметрів. У разі введення некоректних даних система видає повідомлення про помилку та необхідність повторного введення даних.

Далі описано структуру вікон інтерфейсу та їх можливості.

Формування вибірки. У вікні «Формування вибірки» (рис. 5) початкових даних для користувача передбачено можливість завантажити власний підготовлений файл, натиснувши кнопку «Вибрати файл» з тестовими або ре-



Рис. 4. Використовувані режими моделювання

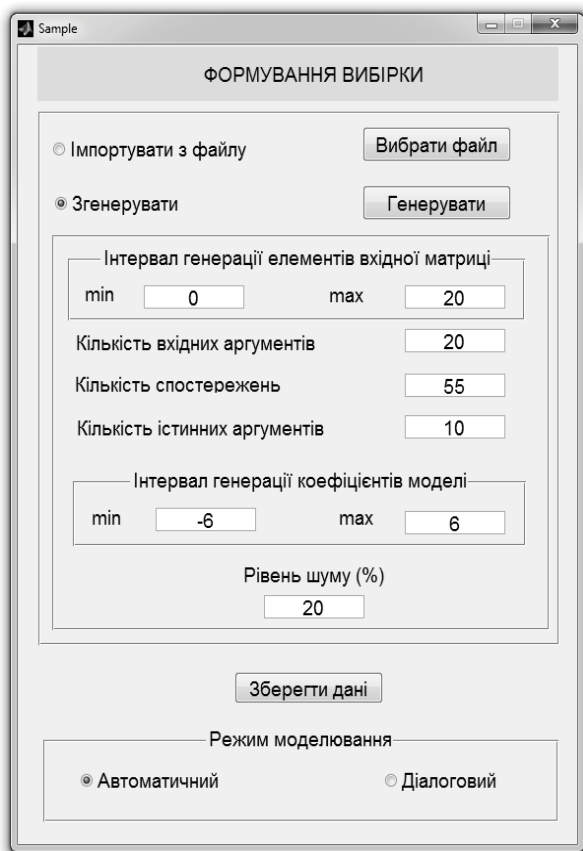


Рис. 5. Вікно «Формування вибірки»

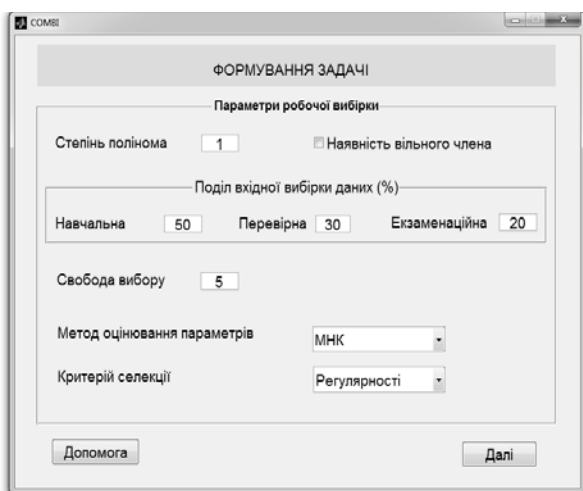


Рис. 6. Вікно «Формування задачі/Параметри робочої вибірки»

альними даними або згенерувати тестові дані безпосередньо у вікні, натиснувши кнопку «Генерувати».

При генеруванні даних для тестової задачі потрібно вибрати інтервал мінімальних і максимальних значень, у якому генеруватимуться елементи початкової матриці аргументів, і задати кількості вхідних аргументів, спостережень та істинних аргументів у моделі.

Потрібно також вказати інтервал, в якому будуть згенеровані коефіцієнти при істинних аргументах тестової моделі та задати рівень шуму у відсотках (за замовчуванням нуль).

Сформовану тестову вибірку початкових даних можна зберегти у файл, натиснувши кнопку «Зберегти». У цьому ж вікні вибирається також режим роботи програмного комплексу — автоматичний або діалоговий.

При виборі режиму моделювання відкривається наступне вікно «Формування задачі/Параметри робочої вибірки» (рис. 6).

У поле «Степінь полінома» цього вікна вводиться натуральне число, що задає клас поліномів, серед яких шукається оптимальна модель. В інші поля задаються: критерій селекції, за яким відбиратимуться кращі моделі (за замовчуванням — це критерій регулярності); метод оцінювання параметрів (за замовчуванням використовується МНК); значення свободи вибору (за замовчуванням п'ять), а також поділ вибірки на три частини (у відсотках): навчальну, перевірну та екзаменаційну.

Навчальна підвибірка потрібна для оцінювання коефіцієнтів структур моделі, перевірна — для вибору кращих моделей за вказаним критерієм селекції, екзаменаційна — для перевірки точності моделі на нових даних, які не брали участі в її побудові.

У разі, коли у відповідних полях нічого не вказувати, за замовчуванням вибірка буде поділена у пропорції 50 : 30 : 20 відповідно.

Після того, як всі поля на вкладці «Формування задачі/Параметри робочої вибірки» будуть заповнені і натиснута кнопка «Далі», відбувається перехід у вікно «Формування задачі/Параметри генетичного пошуку»

(рис. 7), де користувачу необхідно задати відповідні значення.

Розмір початкової популяції. Задається кількість випадково згенерованих структур моделей (послідовностей заданої довжини із одиниць та нулів, двійкових хромосом), серед яких почнеться пошук оптимальної моделі.

Розподіл генерації структур моделей. У цьому полі з випадаючого меню із множини вибирається функція розподілу, яку слід використовувати при генерації початкової популяції.

Розмір проміжної популяції. Задається розмір проміжної популяції, яка складається з особин (структур моделей) з кращими значеннями цільової функції, які будуть використані для формування нащадків. Якісний склад цієї популяції визначається оператором селекції.

Кількість пар для кросинговеру. Цей показник визначає кількість пар хромосом з проміжної популяції, на які може діяти оператор кросинговеру.

Оператор кросинговеру. У цьому полі з випадаючого меню «Оператор кросинговеру» користувач має можливість обрати один з варіантів схрещування: одноточковий, двоточковий, рівномірний.

Імовірність кросинговеру. У цьому полі задається число $p_c \in (0,5, 1)$ — імовірність, з якою до кожної пари хромосом застосовується оператор кросинговеру. За замовчуванням $p_c = 0,9$.

Оператор мутації. У цьому полі з випадаючого меню «Оператор мутації» користувач може обрати один із таких операторів мутації: побітової, нуль-бітової, однібітової.

Імовірність мутації. У цьому полі задається число $p_m \in (0, 0.2]$ — імовірність, з якою до будь-якої хромосоми застосовується оператор мутації. За замовчуванням $p_m = 0,1$.

Кількість запусків. Задається кількість запусків алгоритму, необхідних для усереднення різних характеристик його роботи.

Критерій зупинки ГА: користувач може обрати один з трьох варіантів критерію зупинки ГА.

При виборі критерію *Приріст* алгоритм при-

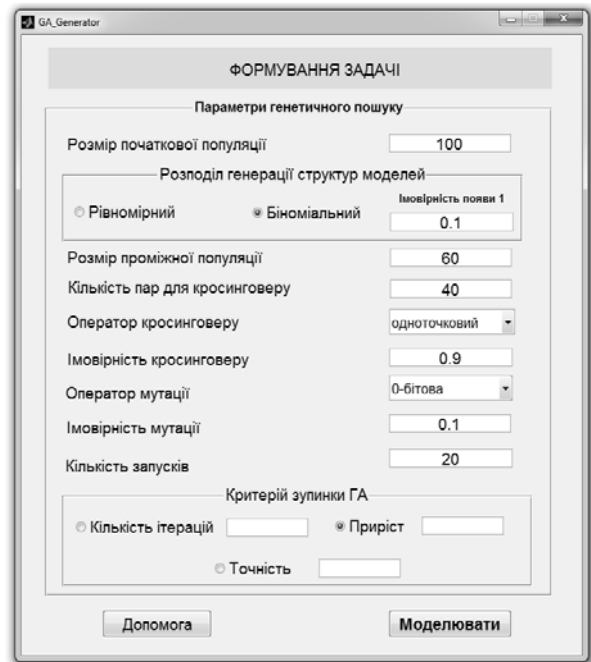


Рис. 7. Вікно «Формування задачі/Параметри генетичного пошуку»

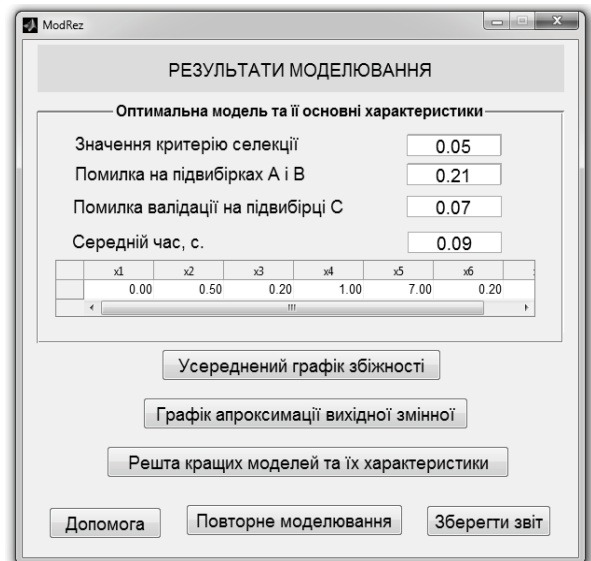


Рис. 8. Вікно «Результати моделювання»

пинить роботу, коли різниця між значеннями зовнішнього критерію селекції кращих моделей на двох сусідніх ітераціях стане менше введеного в поле числа.



Рис. 9. Вікно «Усереднений графік збіжності»



Рис. 10. «Апроксимація вихідної змінної»

	AR_B	MSE_AB	VE_C	x1	x2	x3	x4
1	0.06	0.22	0.08	1.03	0.37	0.71	2.94
2	0.07	0.21	0.09	0.81	0.11	0.49	2.85
3	0.07	0.22	0.09	2.01	0.32	0.74	2.92
4	0.08	0.22	0.11	1.78	0.33	0.71	2.97

Рис. 11. Вікно решти кращих моделей

При виборі критерію *Точність* алгоритм припинить роботу, коли значення зовнішнього критерію селекції досягне вказаної точності.

При виборі критерію *Кількість ітерацій* алгоритм припиняє роботу після здійснення заданої кількості ітерацій.

Після того, як задача сформульована, треба натиснути кнопку «Моделювати» для побудови моделі з заданими характеристиками алгоритму. Після того, як процес моделювання завершено, з’являється вікно «Результати моделювання» (рис. 8).

В цьому вікні спочатку показано оптимальну модель та її основні характеристики — значення зовнішнього критерію селекції та середній час пошуку оптимальної моделі (при проведенні серії запусків), які дають уявлення про якість знайденої моделі.

При натисканні на кнопку «Усереднений графік збіжності» відкривається вікно (рис. 9), в якому зображується усереднений графік збіжності. Користувач за бажанням може зберегти графік у звіт та завантажити його, натиснувши на кнопку «Додати у звіт».

При натисканні на кнопку «Графік апроксимації вихідної змінної» (див. рис. 8) відкривається аналогічне вікно (рис. 10), в якому є цей графік і помилки на підвбірках А, В та С. Користувач за бажання може додати його у звіт, натиснувши кнопку «Додати у звіт».

Кнопка «Решта кращих моделей та їх характеристики» (див. рис.8) активує вікно, в якому представлено характеристики решти кращих моделей (рис. 11).

При натисканні на кнопку «Повторне моделювання» (див. рис. 8) здійснюється перехід знову на вікно «Формування вибірки», і користувач за бажання може змінити налаштування параметрів для досягнення кращого результату.

Натиснувши на кнопку «Зберегти звіт», (див. рис. 8) обрані результати зберігаються у файл звіту, який за замовчуванням має назву «Report» у форматі docx. На початку звіту автоматично виводяться всі значення введених параметрів для пошуку моделі, а далі — оптимальна модель та значення критерію якості.

Функціональні можливості розробленої технології

Розроблена комп'ютерна технологія на основі методу КОМБІ-ГА надає користувачу можливість розв'язувати задачі індуктивного моделювання процесів різної природи, а також досліджувати властивості гібридного алгоритму.

При моделюванні користувач може обрати зручний спосіб задання початкових даних, завантажуючи власний файл в одному з названих раніше форматів або генеруючи тестові дані безпосередньо засобами інтерфейсу. Крім того, застосувавши отриману модель доданих екзаменаційної вибірки, можна імітувати поведінку модельованого об'єкта, процесу або системи.

Дослідження властивостей алгоритму охоплює такі можливості:

- налаштування характеристик алгоритму КОМБІ-ГА для максимального урахування специфіки задачі;
- побудова графіка збіжності алгоритму та отримання відповідного значення мінімуму зовнішнього критерію, що характеризують якість роботи алгоритму;
- побудова усередненого графіка збіжності алгоритму після серії запусків ГА, що точніше характеризує роботу алгоритму;
- порівняння результату роботи КОМБІ-ГА, зокрема побудованої оптимальної моделі та часу її пошуку, з перебірними алгоритмами КОМБІ [4], *MULTI* [6] та *LASSO* [7] для додаткової перевірки коректності роботи алгоритму;
- побудова спільного графіка кривих збіжності впродовж серії запусків ГА для візуалізації розкиду цих кривих, тобто визначення стабільності роботи алгоритму.

Тестова перевірка роботи комплексу

Для демонстрації коректності роботи КОМБІ-ГА розглянемо тестовий приклад з 20 вхідними аргументами, 10 з яких — істинні, та порівняємо результати з алгоритмами КОМБІ, *MULTI* та *LASSO*, що містять оптимальну модель, а також час, необхідний для її пошуку. Комбінаторний

Таблиця 2. Порівняння часу пошуку оптимальної моделі — алгоритмами КОМБІ, *MULTI*, *LASSO* та КОМБІ-ГА

Час пошуку оптимальної моделі, с.			
КОМБІ	<i>MULTI</i>	<i>LASSO</i>	КОМБІ-ГА
84,9	0,48	0,53	0,08

алгоритм здійснює перебір усіх моделей і гарантує пошук оптимального результату.

Алгоритм *MULTI* [6] (багатоетапний комбінаторно-селекційний алгоритм) здійснює спрямований скорочений перебір для визначення оптимальної моделі, тобто результату повного перебору, та реалізує процедуру поетапного формування структур і оцінювання параметрів моделей з послідовним збільшенням складності частинних моделей тільки на один аргумент.

Метод *LASSO* [7] призначений для оцінювання коефіцієнтів лінійної регресійної моделі й полягає у введенні обмеження на норму вектора коефіцієнтів моделі, що веде до перетворення в нуль деяких коефіцієнтів моделі, зменшуючи розмірність. Метод дозволяє отримати інтерпретовані моделі, відбираючи ознаки, які мають найбільший вплив на вихідний вектор.

Оптимальні моделі, побудовані всіма чотирма порівнюваними алгоритмами, мають однакову структуру і коефіцієнти. Згідно з табл. 2, оптимальну модель найшвидше знаходить гібридний комбінаторно-генетичний алгоритм КОМБІ-ГА.

Отримані результати показують високу ефективність і коректність роботи алгоритму КОМБІ-ГА та розробленого програмного комплексу.

Висновки

Програмний комплекс, розроблений на основі комбінаторно-генетичного методу МГУА засобами системи *MATLAB*, призначений для автоматизованого супроводу дій користувача при розв'язанні дослідницьких і реальних задач індуктивного моделювання засобами віконного інтерфейсу.

Цей програмний комплекс надає користувачу широких можливостей, зокрема, дозволяє розв'язувати реальні та тестові задачі моделювання, порівнювати результати моделювання, досліджувати характер збіжності, точність та стійкість

алгоритму залежно від введених параметрів для урахування особливостей конкретної задачі. По завершенні роботи програмного комплексу результати моделювання можна зберегти у файл звіту для подальшого аналізу та використання.

СПИСОК ЛІТЕРАТУРИ

1. Мороз О.Г., Переборный алгоритм МГУА с генетическим поиском оптимальной модели. Управляющие системы и машины, Киев, 2016, №6, С. 73—79.
2. Ануфриев И.Е., Смирнов А.Б., Смирнова Е.Н., MATLAB 7. СПб.: БХВ-Петербург, 2005, 1104 с.
3. Степашко В.С. Элементы теории индуктивного моделирования. Стан та перспективи розвитку інформатики в Україні: монографія / Кол. авторів. К.: Наук. думка, 2010, с. 481—496.
4. Ивахненко А.Г., Степашко В.С., Помехоустойчивость моделирования. Киев: Наук. думка, 1985, 216 с.
5. Мороз О.Г., Степашко В.С., Порівняльний аналіз генераторів структур моделей у перебірних алгоритмах МГУА // Індуктивне моделювання складних систем. Зб. наук. праць. Вип. 8, К.: МННЦ ІТС НАНУ, 2016, с. 133—148.
6. Степашко В.С., Костенко Ю.В. Комбинаторно-селекционный алгоритм последовательного поиска модели оптимальной сложности / Праці I Міжнар. конф. з індуктивного моделювання. Т.1, Ч.1, Львів: ДНДІІІ, 2002, с 72—76.
7. Tibshirani R. Regression Shrinkage and Selection via the Lasso // J. of the Royal Statistical Society. Series B (Methodological), 1996, 58, № 1, p 267—28

Надійшла 16.11.2018

REFERENCES

1. Moroz, O.G., 2016. "Sorting-Out the GMDH Algorithm with the Genetic Search of Optimal Model". Upravlausie sistemy i masiny, 6, pp. 73—79. (In Russian).
2. Anufriev, I.E., Smirnov, AB, Smirnova, E.N., 2005. MATLAB 7. SPb.: BHV-Petersburg, 1104 p. (In Russian).
3. Stepashko, V.S., 2010. "Elements of the theory of inductive modeling". The state and prospects of the development of computer science in Ukraine: a monograph. Col. authors Kyiv: Naukova dumka, pp. 481—496. (In Ukrainian).
4. Ivakhnenko, A.G., Stepashko, V.S., 1985. Pomekhoustoychivost' modelirovaniya (Noise Immunity of Modelling), Kiev, Naukova dumka, 216 p. (In Russian.)
5. Moroz, O.G., Stepashko, V.S., 2016. "Comparative analysis of generators of model structures in interpolated algorithms of GMDH". Inductive modeling of complex systems: Coll. sciences works, K.: IRTC ITS NASU, 8, pp. 133—148. (In Ukrainian).
6. Stepashko, VS, Kostenko, Yu.V., 2002. "Combinatorial selection algorithm for sequential search of optimal complexity model". Works of the 1st International Conference on Inductive Modeling. T.1, Ch.1. Lviv: DNDIII, pp. 72—76. (In Russian).
7. Tibshirani, R., 1996. "Regression Shrinkage and Selection via the Lasso". Journal of the Royal Statistical Society. Series B (Methodological), 58 (1), pp. 267—288.

Received 16.11.2018

O.G. Moroz, junior research scientist,
International Research and Training Center for Information Technologies and Systems
of the NAS and MES of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine
olhahryhmoroz@gmail.com

SOFTWARE PACKAGE FOR INDUCTIVE MODELING BASED ON COMBINATORIAL-GENETIC METHOD

Introduction. The combinatorial-genetic method *COMBI-GA* is an effective self-organizing means of inductive modeling of complex linear and nonlinear objects, systems, and processes of various nature. A software package based on *COMBI-GA*

is developed using the programming language *MATLAB*. It is intended both for solving practical modeling problems from observational data under incomplete information about an object and for investigating the capabilities of *COMBI-GA*, in particular, for the problems of large dimensionality.

The purpose is to describe the structure, interface and functionality of the software package for the inductive construction of the optimal model objects based on the combinatorial-genetic method.

Results. The paper presents a formal description and main steps of the hybrid algorithm *COMBI-GA*. The characteristics and capabilities of the software package based on this algorithm are considered in detail, particularly the general structure, working mechanism, automatic and dialog modes of the modeling process, user interface, qualitative and quantitative indicators of the simulation results etc. The analysis of the *COMBI-GA* effectiveness in the sense of the restoration accuracy of a given test model and the time to find it was carried out in comparison with the *LASSO* algorithm as well as with the sorting-out algorithms *COMBI* and *MULTI*. It is shown that all algorithms find the correct model but *COMBI-GA* does it much faster.

Conclusions. The developed software package provides the user with the wide opportunities, in particular, it allows solving real and testing simulation problems, comparing simulation results, investigating the convergence pattern, accuracy and stability of the algorithm depending on the entered parameters to take into account the features of a specific task.

Keywords: *GMDH, COMBI algorithm, genetic algorithm, hybrid algorithm COMBI-GA, software package.*

О.Г. Мороз, мл. науч. сотр.,

Международный научно-учебный центр информационных технологий и систем
НАН и МОН Украины, просп. Глушкова, 40, Киев 03187, Украина,
olhahryhmoroz@gmail.com

ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ ИНДУКТИВНОГО МОДЕЛИРОВАНИЯ НА ОСНОВЕ КОМБИНАТОРНО-ГЕНЕТИЧЕСКОГО МЕТОДА

Введение. Комбинаторно-генетический метод КОМБИ-ГА является эффективным самоорганизующимся средством индуктивного моделирования сложных линейных и нелинейных объектов, систем, процессов и явлений различной природы. На его основе, средствами языка программирования *MATLAB*, разработан программный комплекс, предназначенный как для решения практических задач моделирования по данным наблюдений в условиях неполноты информации об объекте, так и для исследования возможностей КОМБИ-ГА, в частности, в задачах большой размерности.

Цель статьи — дать описание структуры, интерфейса и функциональных возможностей программного комплекса для индуктивного построения оптимальной модели сложных объектов на основе комбинаторно-генетического метода.

Результаты. Приведены формальное описание и основные шаги работы гибридного алгоритма КОМБИ-ГА. Рассмотрены характеристики и возможности программного комплекса, построенного на основе этого алгоритма: общая структура, механизм работы, автоматический и диалоговый режимы процесса моделирования, пользовательский интерфейс, качественные и количественные показатели результатов моделирования и пр.

Проведен анализ эффективности работы КОМБИ-ГА в смысле точности восстановления заданной тестовой модели и времени ее поиска в сравнении с алгоритмом *LASSO*, а также переборными алгоритмами КОМБИ и *MULTI*. Показано, что все алгоритмы находят правильную модель, однако КОМБИ-ГА выполняет это значительно быстрее.

Выводы. Разработанный программный комплекс предоставляет пользователю широкие возможности, в частности, позволяет решать реальные и тестовые задачи моделирования, сравнивать эти результаты, исследовать характер сходимости, точность и устойчивость алгоритма в зависимости от введенных параметров для учета особенностей конкретной задачи.

Ключевые слова: *МГУА, алгоритм КОМБИ, генетический алгоритм, гибридный алгоритм КОМБИ-ГА, программный комплекс.*