Інструментальний комплекс для дослідження методів емпіричного моделювання

Єфіменко С.М.
Міжнародний науково-навчальний центр інформаційних технологій і систем НАН та МОН України, 03680, Київ, просп. Академіка Глушкова, 40;
syefim@ukr.net

Вступ

Прийняття рішень у різних сферах людської діяльності вимагає засобів, що здатні отримувати точні моделі на основі прогнозу процесів. Складні проблеми, що виникають при моделюванні, пов'язані з великим числом змінних, малою кількістю спостережень і невідомими динамічними зв'язками між змінними. Вони характеризуються:

– недостатньою апріорною інформацією;
– великою кількістю параметрів;
– зашумленими або короткими вибірками даних.

В наш час існують різноманітні методи моделювання за даними спостережень, однак їх ефективність є недостатньо дослідженою. Успішне розв'язання задачі моделювання складних об'єктів, процесів і систем за даними спостережень значною мірою залежить від знання особливостей цих методів.

У зв'язку з цим актуальною є проблема розробки інструментального комплексу для дослідження методів емпіричного моделювання.

Такий імітаційний комплекс дав би можливість:

– самостійно конструювати методи моделювання за даними спостережень;
– порівнювати наявні методи за певними критеріями;
– тестувати різні методи моделювання та їхні компоненти;
– розробляти методики і планувати статистичні випробування;
– проводити імітаційні експерименти (екстраполяцію, прогнозування) з моделями, побудованими за різними методами моделювання;
– поповнювати свої знання про методи моделювання під час роботи з імітаційним комплексом.

Виходячи з вимог до імітаційного комплексу, зазначимо, що це має бути досить складний програмний продукт, створення якого неможливе без ґрунтовних знань предметної області, тому далі зупинимось на особливостях методів моделювання за даними спостережень.

Аналіз структури процесу і методів моделювання

У загальному випадку процес розв'язання задачі моделювання за даними спостережень складається з таких основних етапів [1]:

1. Задання вибірки даних (отриманої в результаті пасивного або активного експерименту), а також апріорної інформації;
2. Вибір або задання класу базисних функцій і відповідного перетворювання даних;
3. Генерація різних структур моделей у вибраному класі;
4. Оцінювання параметрів структур, що генеруються, і формування множини F;
5. Вибір критерію якості CR(f) кожної згенерованої моделі;
6. Мінімізація заданого критерію і вибір оптимальної моделі f;
7. Перевірка адекватності отриманої оптимальної моделі;
8. Застосування моделі (її використання).

Кожен метод моделювання можна описати за допомогою чотирьох компонентів (рис. 1), з урахуванням яких можна порівнювати функціональні особливості заданої множини методів [1]. З іншого боку, для певної множини методів, розбивши кожен з них на такі складові, можна утворити множину класів моделей, генераторів структур, методів оцінювання параметрів і критеріїв відбору моделей і аналізувати ці множини як самостійні об'єкти. Таке розбиття дає можливість сформувати функціональне наповнення програмної системи.

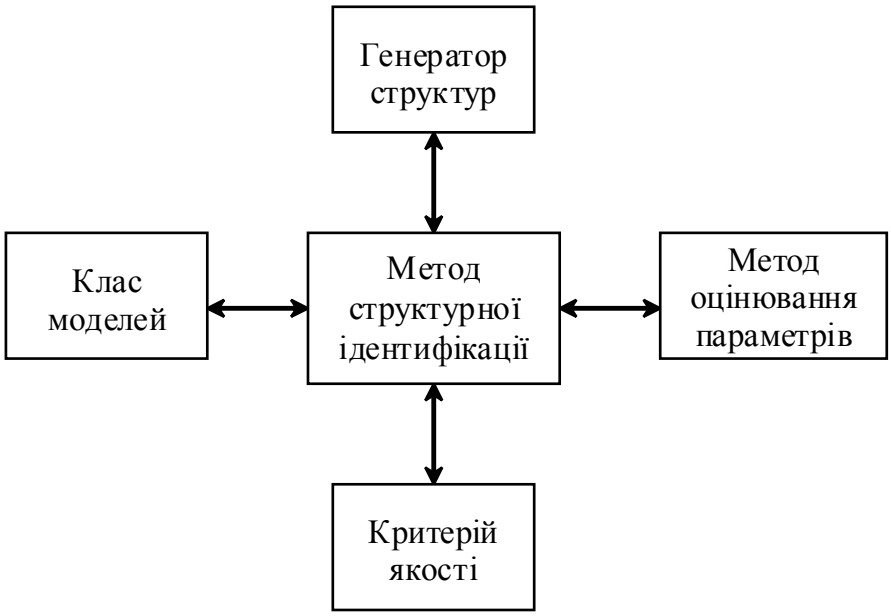

Рис. 1. Компоненти методу побудови моделей

Коротко розглянемо ці компоненти методів моделювання.

В практиці моделювання зустрічаються такі типи задач [1]:

– побудова регресійних моделей статичних об'єктів (у якості базисних функцій застосовуються перш за все поліноми);

– моделювання часових рядів або процесів (застосовуються поліноміальні, тригонометричні, експоненціальні, логарифмічні та логістичні базисні функції);

– моделювання динамічних об’єктів, процесів та систем (перш за все в класі лінійних динамічних моделей).

Генератори структур поділяються на дві групи – перебірні та ітераційні, що відрізняються способами генерації структур моделей та організацією пошуку мінімуму заданого критерію.

Критерії якості, які використовуються при виборі оптимальної моделі, можна розділити на три групи [2]:

– статистичні критерії, що базуються на обчисленні величини помилки моделі на одній і тій же частині вибірки, де й оцінюються параметри, та застосуванні апарату перевірки статистичних гіпотез;

– зовнішні критерії (застосовуються в МГУА) ґрунтуються на розбитті вибірки на дві й більше частин. На одній частині – „внутрішній” – здійснюється оцінювання параметрів, на іншій – „зовнішній” – визначається прогнозна здатність моделей. Застосовуються чотири групи критеріїв: точності, узгодженості, балансу і динамічні;

– критерії з явним штрафом за складність (зокрема, критерії Акаіке та Малоуза) до помилки моделі „додають” адитивні або мультиплікативні штрафні члени, які залежать від кількості точок у вибірці і від складності (числа параметрів) моделі [3, 4].

Всі методи оцінювання параметрів зводяться до прямих (або, інакше, точних) та ітераційних методів [5]. Якщо для оцінювання параметрів застосовується метод найменших квадратів (МНК), тоді задача зводиться до розв’язання систем лінійних рівнянь, для чого можна використовувати різні методи. При цьому точними вважаються методи, які дають розв’язання задачі за допомогою скінченної кількості елементарних арифметичних операцій. До точних методів належать метод Гауса (та його модифікації – методи Гауса-Жордана, Краута), метод квадратних коренів, метод ортогоналізації Грама-Шмідта, метод відображень та інші. Ітераційні методи дають наближений розв’язок. При цьому розв’язок отримуємо як границю послідовних наближень, що обчислюються деяким одноманітним процесом.

Загальна структура імітаційного комплексу

Послідовність усіх етапів розв’язання задачі за допомогою імітаційного комплексу та використовувані при цьому компоненти методів можна подати таким чином (рис.2).

Як випливає з рис.2, імітаційний комплекс як програмний продукт матиме модульну структуру, адже відомо, що складні задачі найкраще виконуються системою взаємодіючих модульних програм, кожна з яких виконує одне завдання і має поєднуватися з іншими компонентами.

Приклади тестувань

Приклад 1. За допомогою запрограмованого модуля оцінювання параметрів покажемо процес розв’язання конкретної задачі – тестування рекурентних методів розв’язування систем лінійних рівнянь для задачі

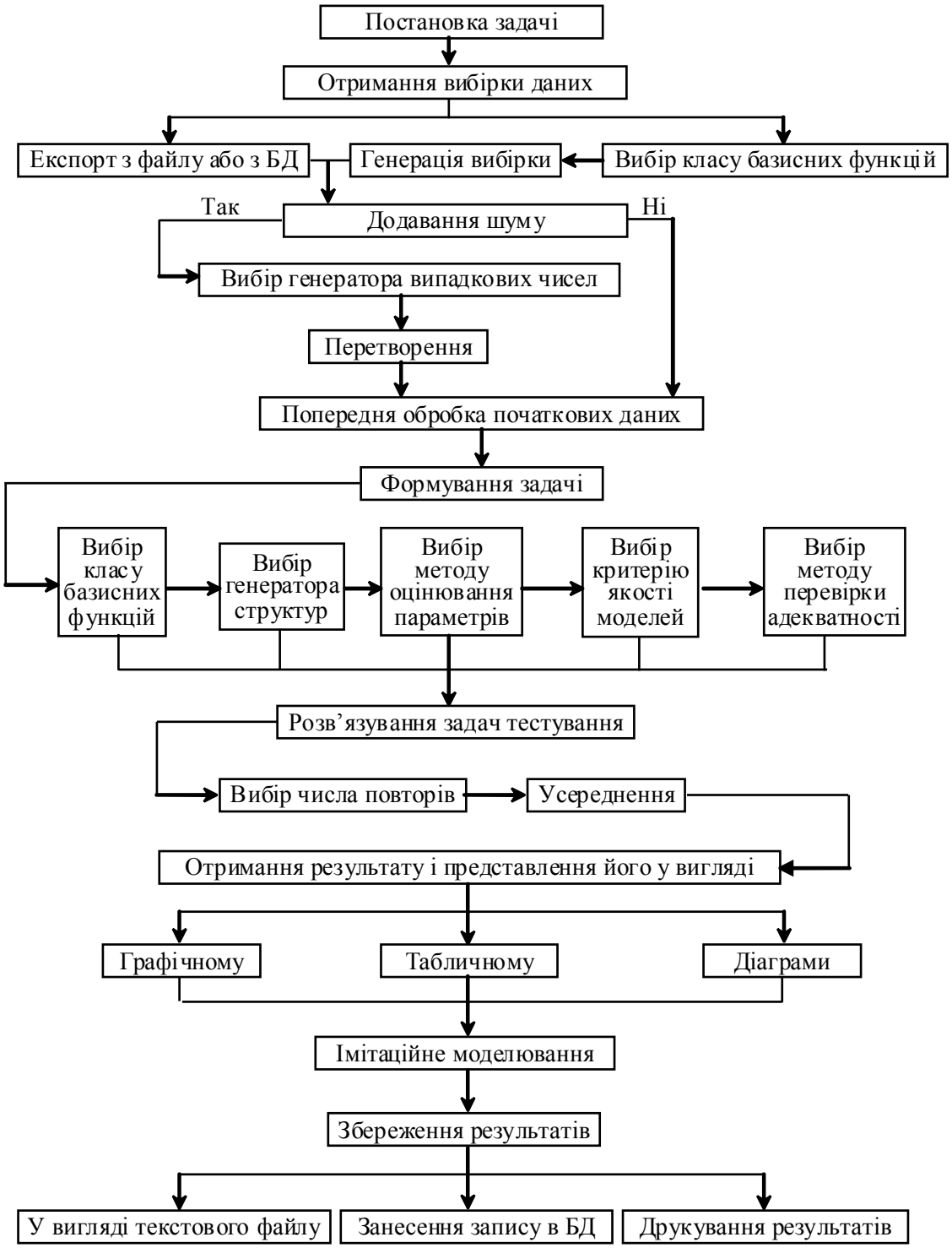

Рис. 2. Послідовність етапів розв'язання задачі за допомогою імітаційного комплексу

оцінювання параметрів: 1) рекурентної модифікації методу Грама-Шмідта; 2) методу обрамлення; 3) рекурентної модифікації методу Гауса. Матриця X містила 300 рівнянь і 280 невідомих. Тестування виконувалось у класі вкладених структур, тобто спочатку оцінювались параметри для моделі, яка містила перший аргумент, потім перших два і так далі до повного включення всіх аргументів. Результати представлені на рис. 3.

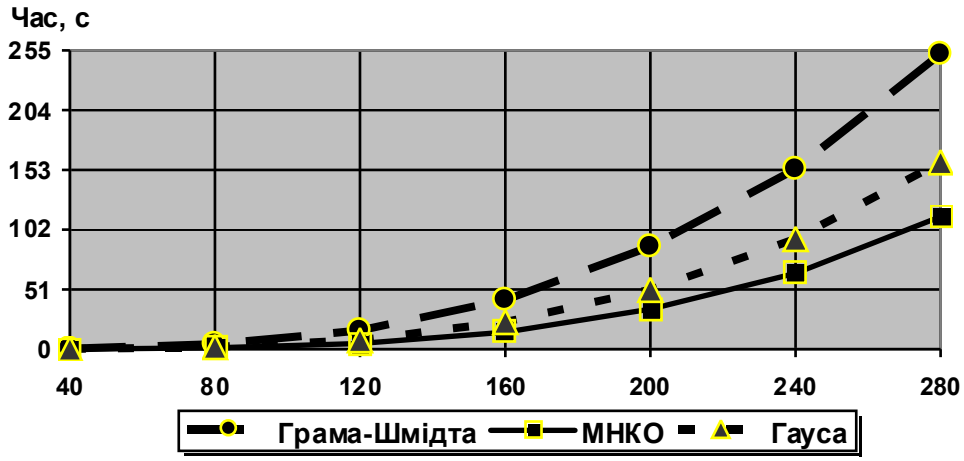

Рис. 3. Тестування рекурентних методів розв'язування систем лінійних рівнянь

Як видно з рис. 3, алгоритм обрамлення, працює найшвидше у класі вкладених структур, однак не є таким же чисельно стійким, як алгоритми Гауса та Грама-Шмідта.

Приклад 2. Тестування, аналогічне попередньому, проведено для класичного алгоритму Гауса та його рекурентної модифікації. Результати, представлені на рис. 4, підтверджують теоретичні висновки про те, що трудомісткість розрахунку коефіцієнтів, тобто кількість елементарних арифметичних операцій, при включенні в модель додаткового аргумента пропорційна третьому степеню складності моделі при використанні нерекурентних алгоритмів і другому степеню – при використанні рекурентних.

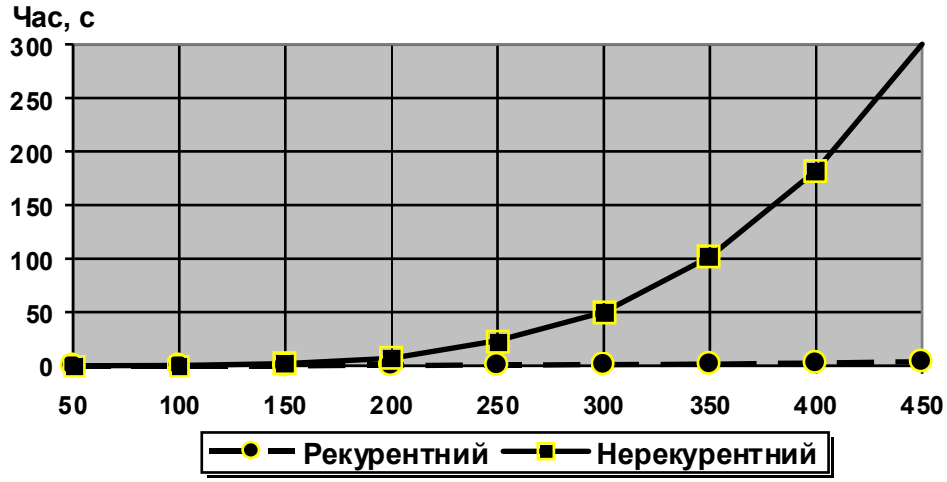

Рис. 4. Тестування рекурентного і нерекурентного алгоритмів Гауса

Приклад 3. Виконано тестування трьох критеріїв якості вибору моделей: 1) критерію регулярності $Ar(s) = \left\| y_B - X_B \hat{\theta}_{As} \right\|^2$, де A і B – відповідно навчальна і перевірочна частини вибірки; 2) критерію Малоуза $Cp(s) = RSS(s) + 2\sigma^2 s$; 3) критерію Акаіке $FPE(s) = \frac{n+s}{n-s} RSS(s)$, де $RSS(s) = \left\| y - \hat{y}_s \right\|^2$, а n – кількість рівнянь. Результати представлено на рис. 5.

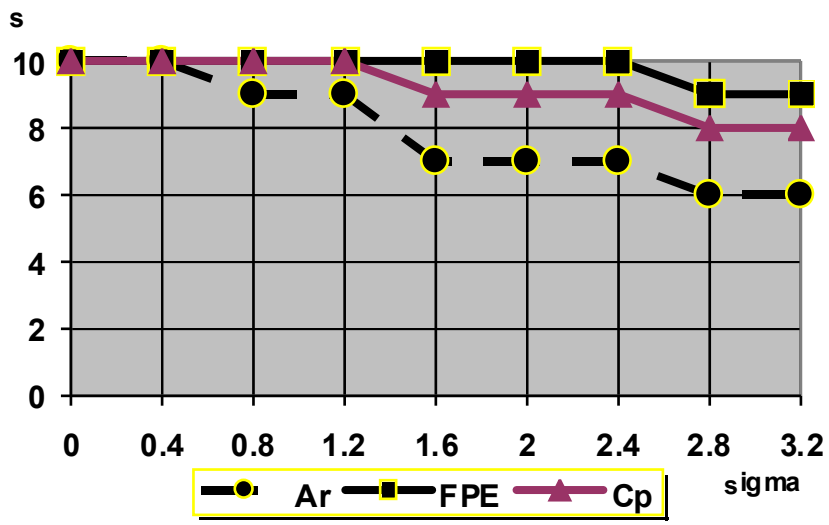

Рис. 5. Тестування критеріїв якості вибору моделей

Тестування проводилось таким чином: Генерувалась матриця X, що містила 15 аргументів і 30 рівнянь. Вектор y формувався як лінійна комбінація перших 10 аргументів (тобто решта 5 були зайвими) з додаванням згенерованого шуму. По вертикальній осі відкладена складність оптимальної моделі, отриманої за мінімумом відповідного критерію. По горизонтальній осі – значення дисперсії шуму. У даному експерименті критерій Малоуза виступає у якості ідеального критерію, адже у обчислювальній формулі містить істинне значення дисперсії шуму. Графік даного критерію показує, як має спрощуватись модель із зростанням шуму. Згідно з результатами даного тестування критерій регулярності можна вважати ефективним, оскільки він, на відміну від критерію Акаіке, не переускладнює модель із зростанням шуму.

Висновки

Розроблено структуру комплексу інструментальних засобів для тестування методів моделювання за експериментальними даними. Наведено результати тестування окремих компонентів методів моделювання та експериментально показано, що застосування рекурентних методів розв'язування систем лінійних рівнянь для задачі оцінювання параметрів істотно зменшує обчислювальні затрати при моделюванні.

Література

1. Степашко В.С. Алгоритмы МГУА как основа автоматизации процесса моделирования по экспериментальным данным // Автоматика. - 1988. - № 4. - С.44-55.
2. Степашко В.С., Кочерга Ю.Л. Методы и критерии решения задач структурной идентификации // Автоматика. - 1985. - №5. - С.29-37.
3. Akaike H. Fitting autoregressive model for prediction // Ann. Stat.- 1969.- v.21.- P.243-247.
4. Mallows C.L. Some comments on Cp // Technometrics.- 1973.- V.15.- P.661-667.
5. Фаддеев Д.К., Фаддеева В.Н. Вычислительные методы линейной алгебры. Изд. 2-е.- М.-Л.: Физматгиз, 1963.