

С.О. Резнікова, ІПМЕ ім.Г.Є.Пухова НАН України, Київ
М.Ю. Савченко, ІПМЕ ім.Г.Є.Пухова НАН України, Київ
О.А. Чемерис, к.т.н, ІПМЕ ім.Г.Є.Пухова НАН України, Київ

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ВИКОНАННЯ ПРОГРАМ З ПАКЕТУ rARMS ДЛЯ РОЗПОДІЛЕНИХ СИСТЕМ

Так як кожен обчислювальний кластер має свої унікальні характеристики, а задачі, які виконуються на ньому, також значно відрізняються одна від одної, то визначення ефективності обробки даних є достатньо складним процесом, так як на показники ефективності впливають як сам алгоритм програми, так і характеристики конкретної обчислювальної системи, а також їх відповідність одне одному. В загальному, для отримання відомостей про ефективність виконання програм необхідно виконати цілий спектр досліджень, серед яких а) аналіз конфігурації кластеру; б) аналіз ефективності системного та програмного забезпечення; в) аналіз структури програми; г) аналіз алгоритму.

Відомі методи, які використовуються у дослідженнях, не здатні достатньо швидко та універсально оцінити ефективність виконання програми на конкретній обчислювальній системі, а також не мають функцій прогнозування часу виконання паралельних програм. Зокрема програмні аналізатори часто є платформи-орієнтованими, вимагають включення своїх модулів (чи директив) у програмний код, що значно звужує їх використання. Виходячи з позиції, що ефективного методу суміщеної оцінки немає, тому і задача оцінки ефективності виконання програм на розподілених обчислювальних системах є задачею актуальною.

Розв'язанням даної проблеми може бути використання статистичних методів побудови моделей, які здатні прогнозувати продуктивність обчислень або час виконання програми, не запускаючи при цьому саму програму з конкретними даними. Такі моделі, що можуть бути доданими до програмного забезпечення, допоможуть зробити попередні оцінки ефективності виконання програми на конкретній обчислювальній системі і підібрати параметри задачі для отримання максимальної швидкодії для конкретного кластеру та операційної системи (ОС).

Тому для розробки моделі ефективності програм, а саме обробки розрізаних масивів даних, описаних в [1], був вибраний статистичний підхід на основі регресійних методів, в якому для розв'язання вищезазначених задач в якості вхідних даних можна використовувати велику кількість конкретних параметрів. При цьому отримуємо оцінку, при яких параметрах алгоритму його виконання буде найбільш ефективним.

Загальна схема побудови та використання статистичних моделей для оцінки ефективності виконання паралельних програм складається з декількох

етапів:

1) *Визначення цілей моделювання.* Створені моделі повинні описувати поведінку паралельної програми на даній обчислювальній системі. Можуть бути дані попередні оцінки ефективності виконання паралельної програми. Модель, що описує поведінку системи можна використовувати для визначення числа процесорів системи, коли програма виконується найбільш ефективно. Крім того, може виявитися, що час на синхронізацію процесів, передачу даних, доступ до загальних ресурсів і т.д. виявляється більше часу виконання послідовної програми. У цьому випадку немає необхідності в паралельних обчислювальних системах.

2) *Специфікація елементів системи.* Починаючи процес побудови моделі необхідно визначити середовище, в якому функціонує програма. Необхідно виділити ті змінні, які мають найбільший вплив на якість моделювання. Для цього потрібно отримати безліч статистичних вимірювань, причому, чим більша кількість вимірів, тим точніше може бути модель. Отриманий статистичний матеріал повинен бути класифікований, коли виділяють а) змінні незалежні (головні і допоміжні), що впливають на залежні змінні; б) залежні змінні – описують наслідки зміни залежних змінних; в) перешкоди, контекстні змінні, керуючі та інші.

Вихідною змінною є або час виконання програми, або прискорення, або ефективність виконання програми на паралельній обчислювальній системі. В якості незалежних змінних виступають число процесорів, кількість потоків у програмі, розмірність потоків, розмір оперативної пам'яті і розмір дискового простору і т.д. Деякі параметри не можуть бути використані в моделях, тому що вони закладаються в процесі проектування програми і на них немає ніякого впливу. До таких параметрів відносяться метод і алгоритм розпаралелювання, параметри операційної системи і т.д.

Таким чином, потрібно визначити статистичний зв'язок вихідної змінної від інших змінних. На основі набору статистичних даних потрібно довести логічний зв'язок між змінними та визначити які змінні чинять найбільший вплив на вихідну змінну.

3) *Вибір класу моделі.* Після етапу визначення статистичних даних необхідно цей матеріал систематизувати відповідним чином, щоб перейти від значень індивідуальних до простору пов'язаних значень. Відповідно до аналізу даних слід визначити вид регресії. Для цього можна скористатися лініями тренда, які покажуть загальний вигляд даних і функцію апроксимації. Слід враховувати, що лінія регресії повинна характеризуватися R^2 якомога близьким до одиниці і якомога меншим значенням статистики Маллоуза S_r [2]. Функція регресії може бути однією з наступних а) лінійна; б) степеневою; в) логарифмічною; г) показниковою; д) визначеною користувачем.

При неправильному виборі функції апроксимації необхідно знову провести дослідження для іншого виду функції, дослідивши при цьому, як мінімум, R^2 .

4) *Естимация коэффициентов и верификация модели.* У процесі естимачії визначаються коефіцієнти змінних. Чим більше використовується різних методів оцінки, то тим більш правдоподібною є модель. Для естимачії використовуються тести і статистики, як описані вище, так і безліч додаткових тестів, таких як, наприклад, а) стандартна похибка β , яка показує значення похибки параметра регресії у разі використання його як незалежної змінної; б) статистичний рівень відповідності r коефіцієнтів регресії; в) коефіцієнт кореляції різноманітності – показує рівень залежності вихідної змінної від інших змінних моделі; г) коефіцієнт детермінації показує, як модель визначає формування вихідної змінної.

При верифікації моделі на основі довільної вибірки із статистики незалежних змінних отримуємо значення вихідної змінної. Якість моделі показує співпадання значень, отриманих на основі моделі і статистичних даних.

5) *Використання моделі для прогнозування.* Це є останній етап процесу моделювання. Він полягає в формулюванні висновків при узагальненні результатів, отриманих на випадковій вибірці на всьому діапазоні значень змінних і ухваленням рішення про параметри при виконанні правил, розроблених на основі певного значення ймовірності. Основним завданням при цьому є можливість визначення значення одержуваних у процесі моделювання помилок.

Дослідження згідно з описаною вище методикою були проведені для програмного пакету pARMS. pARMS (parallel Algebraic Recursive Multilevel Solvers) – це бібліотека розв'язувальних програм для паралельного ітераційного розв'язання великих розріджених систем лінійних рівнянь загального вигляду. Вона базується на попередній обробці даних наближених підобластей Крылова і використовує точку огляду з доменною декомпозицією. Зокрема, пакет використовується для моделювання динамічних систем з зосередженими та розподіленими параметрами, до яких також відносяться об'єкти енергетики. Основна методологія pARMS ґрунтується на рекурсивній багаторівневій факторизації ILU, яка дозволяє розробляти безліч розв'язувальних програм типу доменною декомпозиції в єдиній структурі. Більш детально пакет представлено в [3].

Відповідно до методики було обрано множину вхідних параметрів, які представлено в табл. 1. Слід зазначити, що в результаті проведених експериментів з матрицями різних типів та структур було помічено, що для пакету pARMS найкраще підходять матриці real unsymmetric.

В загальному вигляді вираз моделі часу виконання програм на паралельній обчислювальній системі, як вираз лінійної регресійної моделі, має вигляд

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k + \varepsilon, \quad (1)$$

де b_0, b_1, \dots, b_k – невідомі коефіцієнти (константи), ε – випадкові величини, які не спостерігаються. Для побудови статистичної моделі був

вибраний програмний пакет STATISTICA, а саме модуль *Множинна регресія (Multiple regression)*.

Основна задача побудови моделі полягає в тому, щоб по спостереженням за залежними і незалежними змінними Y та X відповідно, а) оцінити параметри моделі b_0, b_1, \dots, b_k найкращим чином; б) побудувати довірчі інтервали для b_0, b_1, \dots, b_k ; в) перевірити гіпотезу про значимість регресії; г) оцінити степінь адекватності моделі і т.д. Оцінюванням параметрів моделі є побудова прямої методом найменших квадратів, рівняння якої і є рівнянням регресії.

Таблиця 1

Перелік вхідних параметрів для побудови статистичної моделі

Вхідні параметри	Значення
кількість процесорів	1 – 8
передобумовлювач	add_arms, add_ilut, sch_sgs, sch_gilu0
допустимі відхилення для вхідних та вихідних ітерацій (<i>tol</i>)	0.000000001 - 0.001 (кроками *10)
розмір підпростору Крилова для вхідних ітерацій	1, 2, 3, 4, 5
MCA Mpi_leave_pinned	0, 1
обмеження максимальної кількості вихідних ітерацій	1000

Для отримання нормальних залишків та стабілізації дисперсії було застосовано деякі перетворення незалежних змінних, а саме логарифмічне перетворення та добування квадратного кореня, в результаті чого були додані наступні вхідні параметри такі, як десятковий логарифм від допустимих відхилень для вхідних та вихідних ітерацій ($\log_{10}(tol)$), корінь квадратний та квадрат кількості процесорів.

В результаті було отримано таку модель часу виконання програми з пакету pARMS

$$Y = -18.1732x_1 - 99.8708x_2 - 18.3916x_3 - 10.4706x_4 + 172.231x_5 + 5.0438x_6, \quad (2)$$

де фігурують такі вхідні змінні: $x_1 - \log_{10}(tol)$; $x_2 -$ число процесорів; $x_3 -$ передобумовлювач; $x_4 -$ розмір підпространства Крилова; $x_5 - \sqrt{\text{кількість процесорів}}$; $x_6 -$ (кількість процесорів)².

Графік залежності отриманих експериментальних даних від залишків, показаний на рис. 1, і доводить достатню адекватність побудованої моделі, тобто більшість даних лежать або на лінії регресії, або в межах отриманого показника R^2

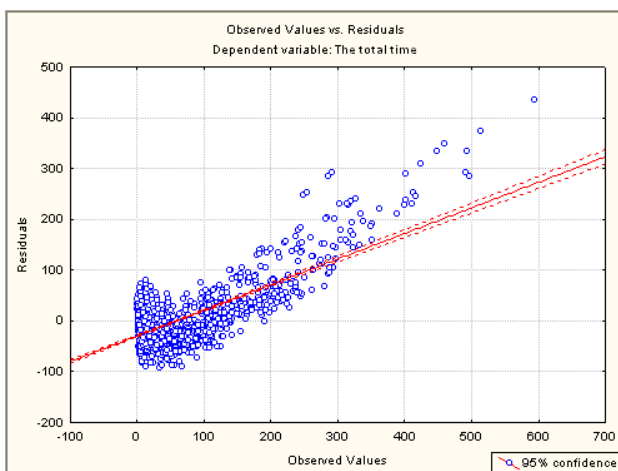


Рис. 1. Залежність експериментальних даних від залишків

На рис. 2 показаний графік вірогідності нормального розподілення побудованої моделі.

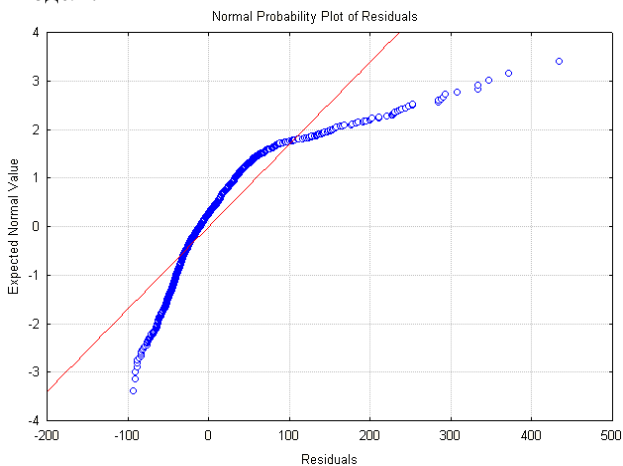


Рис. 2. Графік вірогідності нормального розподілення моделі

Отже, побудовано статистичну модель залежності часу виконання паралельних програм в пакеті rARMS від набору вхідних даних, яку можна використати без запуску задач на всьому діапазоні даних (а такі задачі можуть бути дуже складними та трудомісткими, вимагати значний машинний час) для отримання прогнозного часу їх виконання.

Тепер перевіримо адекватність побудованої моделі (2), порівнявши результат роботи моделі та реальні значення. Діаграма порівняння часу

виконання розрахунків експериментальних даних і даних, отриманих із використанням статистичної моделі наведена на рис. 3.

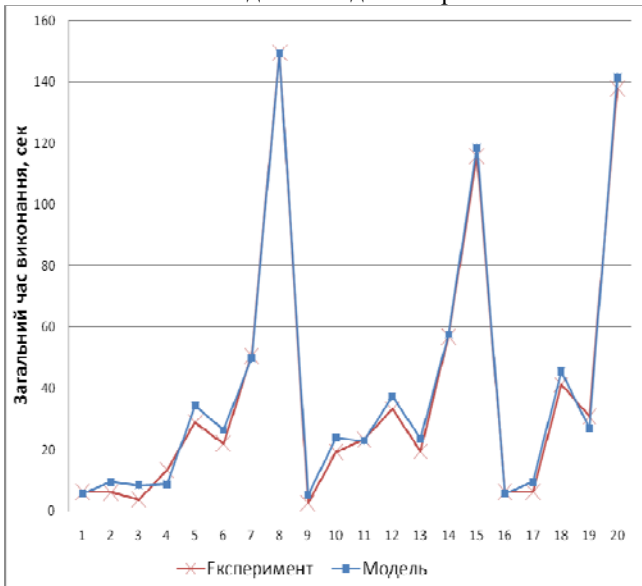


Рис. 3 Порівняння експериментальних даних і даних, отриманих із використанням статистичної моделі

Таким чином, використання статистичних методів для аналізу часу виконання алгоритмів на паралельних комп'ютерах є актуальним завданням. При цьому можуть бути отримані відповіді як на питання ефективності виконання алгоритмів на конкретній обчислювальній системі, так і на питання відповідності виконуваного алгоритму і даної паралельної ОС. Основними перевагами застосування такого підходу є 1) можливість прогнозування ефективності виконання програми без запуску на реальних даних, що вимагає великих затрат, як тимчасових, так і ресурсів; для побудови статистичних моделей достатньо лише знання досить зручних стандартних статистичних пакетів з дружнім інтерфейсом; 2) можливість полегшити аналіз ефективності для задач, які використовуються дуже часто, або для таких, які мають велику складність і тривалий час виконання; побудова статистичної моделі є досить універсальним засобом аналізу ефективності виконання паралельних програм, який не залежить ні від платформи, ні від використання складних засобів аналізу, які часто мають риси вузької спеціалізації та дорого коштують; 3) використання такої статистичної моделі, доданої в програмне забезпечення, допоможе зробити попередні оцінки ефективності виконання програми на конкретній обчислювальній системі і підібрати параметри задачі для отримання максимальної швидкодії при розв'язанні задачі на цій ОС.

Подальшими кроками можуть бути удосконалення методики побудови статистичних моделей, а також знаходження і дослідження обмежень, які можуть бути наявні при використанні цього методу аналізу.

1. *А.А.Боровков*. Математическая статистика: оценка параметров проверка гипотез – М.: Наука, 1984. – 245 с.
2. *Э. Таненбаум*. Архитектура компьютера, СПб.: Питер, 2002. - 704 с.
3. *Yousef Saad, Masha Sosonkina*, pARMS: A package for the parallel iterative solution of general large sparse linear systems user's guide. Report UMSI2004-8, Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, MN, 2004.

Поступила 8.09.2010р.

УДК 376.1:163.912

Б.В.Дурняк, д.т.н., УАД, Л.С. Сікора, д.т.н., НУ «Львівська політехніка»
М. Поліщук, вч. ЛВПУКТтаБ, Р.А. Федчишин, директор ЛВПУКТтаБ

ЦИКЛІЧНА МОДЕЛЬ АКТИВІЗАЦІЇ КОГНІТИВНИХ РЕСУРСІВ ОСОБИ В ПРОЦЕСІ ПРОФОРІЄНТОВАНОГО НАВЧАННЯ

Анотація. В статті розглянуто концепцію циклічної моделі активізації когнітивних ресурсів особи, використання якої дозволяє підняти рівень підготовки в процесі навчання.

Ключові слова. Модель, активація навчання, когнітивний ресурс, цикл навчання

Актуальність. Сучасний етап розвитку виробничих структур в енергетиці, нафтохімічній та транспортних галузях характеризується насиченістю його автоматизованими системами управління [1-5]. При ієрархічній структурі стратегічного і оперативного управління виробництвами для кожного рівня є певні особливості діяльності відповідно на нижніх рівнях необхідні знання фізичних і енергетичних процесів, на верхніх знання з інформаційних і управлінських технологій. В умовах екстремальних і кризових ситуацій, граничних режимах потенційно небезпечних об'єктів які входять в ці структури, на всіх рівнях необхідні інтелектуальні компоненти прийняття рішень з врахуванням фізики, термодинаміки і енергетики технологічних енергоактивних об'єктів. При цьому персонал житнього рівня характеризується уміннями профорієнтованої діяльності і недостатнім рівнем знань в області інформаційних технологій, а верхній рівень персоналу має досить високий інтелектуальний потенціал але він відірваний від змісту і знань про фізичні і енергетичні процеси. Тобто в