УДК 004.942

Застосування паралельних обчислень у задачах ідентифікації параметрів технологічних процесів

А.І. Купін[1], І.О. Музика[1]

Abstract – **Possibility of application of parallel calculations in problems of mathematical models identification of static technological objects by regression analysis is considered.**

Keywords – **Regression analysis, Parallel calculations, Multiprocessor computer.**

I. Вступ

Як відомо, якість керування будь-яким технологічним процесом залежить від багатьох чинників: точності математичної моделі, яка застосовувалася при створенні алгоритму керування, рівня стохастичних збурень, ступеня нелінійності та стабільності робочих параметрів об'єкта керування. Тому сьогодні на підприємствах гірничої галузі промисловості все частіше створюються та впроваджуються спеціалізовані системи підтримки прийняття рішень [1]. Такі розробки покликані проводити оптимізацію техніко-економічних показників виробництва, вони об'єднують в єдину систему сучасні засоби інформаційних технологій (ІТ), автоматичні системи керування та людські ресурси.

II. Постановка проблеми

Проводячи аналіз математичної моделі підготовки гірської маси у кар'єрі гірничо-збагачувального комбінату, авторам вдалося вдосконалити залежність сумарних питомих витрат на видобуток 1 т залізної руди від технологічних параметрів буро-вибухових робіт. Виявилося, що для побудови регресійної моделі необхідно врахувати близько 14 предикторних змінних. Поступове включення їх у модель за допомогою кореляційного аналізу, розрахунок коефіцієнта детермінації, аналіз залишків, матричні розрахунки вимагають застосовувати достатньо потужні обчислювальні засоби. Так, наприклад, проводячи аналіз за допомогою лінійної, квадратичної та кубічної регресії при п'яти змінних, потрібно перевірити понад 90 рівнянь, а при 14 змінних – це значення досягає майже 50 тис. рівнянь. Для цього авторами було розроблене спеціальне програмне забезпечення на мові C++ із використанням бібліотеки Win32 API. Тестування роботи програми показало, що для бази даних, яка містить близько 100 записів, на процесорі з тактовою частотою 1 ГГц лінійний регресійний аналіз триває близько 3 с, квадратичний – більше 1 хв. Тому для створення інтелектуальної системи підтримки прийняття рішень, яка працюватиме в режимі часу близькому до реального, необхідно запропонувати підхід до застосування паралельних обчислень у задачах такого типу та оцінити його ефективність.

III. Аналіз публікацій по темі дослідження

Питанням побудови математичних моделей присвячено багато робіт [2, 3]. Головним чином такі моделі ґрунтуються на застосуванні методів групового урахування аргументів, інтерполяції за допомогою сплайнів, нейромережевих структур та ін. Одним із основних завдань при цьому є досягнення необхідної точності моделі, яка залежить від кількості врахованих незалежних змінних та обсягу навчальної вибірки. Існуючі підходи до зменшення розмірності (метод головних компонент, факторний аналіз) є достатньо складними з точки зору обчислювальної потужності. Сьогодні цю задачу можна ефективно розв'язувати за допомогою багатоядерних обчислювальних машин. Тому питання застосування паралельних обчислень у задачах ідентифікації математичних моделей є зараз актуальним та потребує дослідження.

IV. Викладення матеріалу та результати

У якості регресійного рівняння був взятий класичний поліном Колмогорова-Габора (1)

$$y(X) = \beta_0 + \sum_{i=1}^{m}\beta_i x_i + \sum_{i=1;\,j\ge i}^{m}\beta_{ij} x_i x_j + {} + \sum_{i=1;\,j\ge i;\,q\ge i}^{m}\beta_{ijq} x_i x_j x_q \,, \tag{1}$$

де m – кількість змінних; x_1, x_2, …, x_m – незалежні предикторні змінні; β_i, β_{ij}, β_{ijq} – коефіцієнти регресії, одержані методом найменших квадратів. Вектор коефіцієнтів B можна знайти, розв'язавши матричне рівняння (2)

$$M_\Sigma B = V_Y \Rightarrow B = M_\Sigma^{-1} V_Y \,, \tag{2}$$

де M_Σ , V_Y – матриця і вектор елементів, сформовані із сум вигляду $\sum_{t=1}^{N} x_{i_t} x_{j_t} \ldots x_{q_t}$ і $\sum_{t=1}^{N} x_{i_t} x_{j_t} \ldots y_t$ відповідно; N – обсяг навчальної вибірки.

Для визначення обчислювальної складності було оцінено кількість арифметичних операцій, які виконує центральний процесор (CPU) при формуванні вищезазначених матриць та векторів (таб. 1). До уваги не бралися операції записування та зчитування із оперативної пам'яті. Матричне рівняння розв'язувалося за допомогою метода Гауса із прямим та зворотнім ходом. Слід зауважити, що питома вага розрахунків при формуванні регресійної матриці та її розв'язання сильно відрізняються. Так, при обсягу навчальної вибірки 50

[1] Криворізький технічний університет, вул. XXII Партз'їзду 11, Кривий Ріг, 50027, УКРАЇНА, E-mail: musicvano@mail.ru

точок та 8 незалежних змінних у моделі на формування матричного рівняння (2) витрачається в 20 разів більше часу, ніж на його розв'язання.

Таблиця 1

Обчислювальна складність регресійного аналізу

	Приблизна кількість операцій CPU	
Регресійний аналіз	лінійний	квадратичний
Розрахунок матриці M_{Σ}	$2N(m+1)^2$	$(3/4)N(m+1)^2(m+2)^2$
Розрахунок вектора V_Y	$2N(m+1)$	$(3/2)N(m+1)(m+2)$
Розв'язання методом Гауса	$(2/3)(m+1)^3+$ $+O(m^2)$	$(1/12)(m+1)^3(m+2)^3+$ $+O(m^2)$
Загалом	$2N(m+1)(m+2)$ $+$ $+(2/3)(m+1)^3$	$(3/4)N(m+1)^2(m+2)^2$ $+$ $+(3/2)N(m+1)(m+2)+$ $+(1/12)(m+1)^3(m+2)^3$
Асимптотичний результат при $N\to\infty$; $m\to\infty$	$2Nm^2$	$(3/4)Nm^4$

Графіки, зображені на рис. 1, демонструють наскільки сильно зростає обчислювальна складність при переході від лінійного регресійного аналізу до квадратичного.

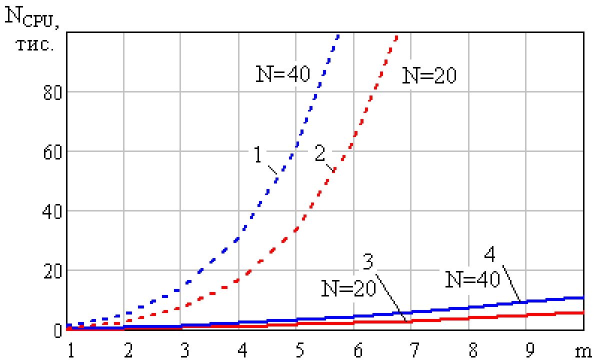

Рис.1. Залежність кількості арифметичних операцій від числа незалежних змінних у моделі у випадку лінійної регресії (3, 4) та квадратичної (1, 2) при різному обсягу навчальної вибірки

З метою прискорення розрахунків за допомогою сучасних обчислювальних засобів було проаналізовано основні підходи паралельної обробки інформації. До них належать такі технології як Application Programming Interface (API), Message Passing Interface (MPI), Parallel Virtual Machine (PVM) та Common Object Request Broker Architecture (CORBA). Проаналізувавши переваги та недоліки цих методів, було вирішено застосувати програмні можливості Win32 API, оскільки для даної технології хоча і характерна відносна складність відлагодження програмних продуктів, проте, працюючи із потоками та процесами операційної системи, вона не потребує спеціалізованих бібліотек.

Згідно з законом Амдала прискорення виконання програми за рахунок розпаралелювання її інструкцій обмежене часом, який потрібен для її послідовних операцій

$$\lambda = \frac{1}{\alpha + (1-\alpha)/p}, \qquad (3)$$

де λ – коефіцієнт, який показує, у скільки разів програма виконується швидше на p процесорах, ніж на одному; α – частина обчислень, яку можна одержати тільки послідовними розрахунками.

На рис. 2 представлена залежність прискорення роботи алгоритму залежно від кількості процесорів. Як видно із графіка, застосування паралельних обчислень тим вигідніше, чим складніша регресійна модель.

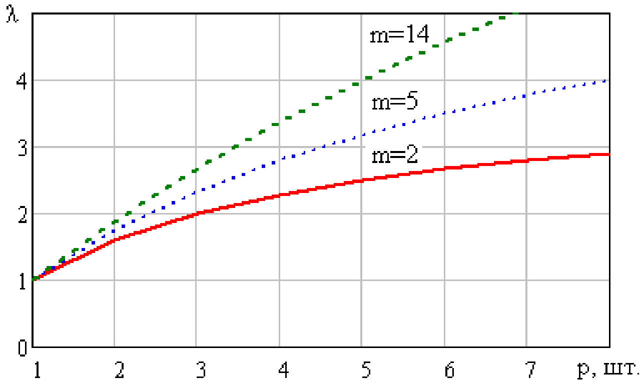

Рис.2. Прискорення роботи алгоритму лінійного регресійного аналізу в залежності від кількості процесорів та числа незалежних змінних у моделі

V. Висновок

Таким чином, застосування паралельних обчислень в системі підтримки прийняття рішень при розв'язанні багатофакторних задач дозволить: прискорити процедуру ідентифікації параметрів технологічних процесів на багатоядерних обчислювальних системах (2-ядерних – не менше ніж у 1,5 рази; 4-ядерних – 2,5-3 рази); проводити квадратичний та кубічний регресійний аналіз у режимі близькому до реального часу. Подальші дослідження авторів будуть спрямовані на створення ефективних паралельних алгоритмів ідентифікації запізнювання за допомогою взаємокореляційних функцій.

Список посилань

[1] Лысенко В.С. Обзор и анализ современных СППР на предприятиях открытой добычи руд / В.С. Лысенко // Зб. наук. праць: Економіка: проблеми теорії та практики. – Дніпропетровськ: ДНУ, 2005. – Вип. 208. – Т. 4. – С. 1091–1098.

[2] Дрейпер Н. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. – М.: Вильямс, 2007. – 912 с.

[3] Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем / Ивахненко А.Г. – К.: Наук. думка, 1981. – 296 с.