УДК 004.272.26

В. Я. Лісовець, Г. Г. Цегелик (Львівський нац. ун-т ім. І. Франка)

КОЕФІЦІЄНТИ ПРИСКОРЕННЯ Й ПАРАЛЕЛЬНОЇ ЕФЕКТИВНОСТІ МЕТОДІВ ПАРАЛЕЛЬНОГО ПОШУКУ

Speedup and parallel coefficients are considered for researching m-parallel sequenced search method and two variant of the method of m-parallel block search. We research the effectiveness of the method for different probability distribution of record request frequency (discrete uniform, binomial, Zipf and generalized the partial occasion of witch is the probability distribution approximately satisfying the rule "80 - 20").

Розглянуто використання коефіцієнтів прискорення та паралельної ефективності для дослідження продуктивності методу m-паралельного послідовного перегляду та двох варіантів методу m-паралельного блочного пошуку. Проведено розрахунок цих коефіцієнтів для паралельних методів пошуку у разі таких законів розподілу ймовірностей звертання до записів, як рівномірний, бінарний, Зіпфа та узагальнений, частковим випадком якого є розподіл, який наближено задовольняє правило "80 - 20".

1. ВСТУП

Завдяки високій надійності та продуктивності багатопроцесорні ЕОМ широко використовують для підтримки й організації великих баз даних (БД). Під час розв'язування різноманітних задач з використанням БД головний акцент переноситься з процедур опрацювання інформації на процедури організації збереження та пошуку інформації в них. Тому продуктивність обчислювальних систем, орієнтованих на роботу з великими БД, здебільшого визначена ефективністю методів пошуку інформації.

Дослідження ефективності методів пошуку інформації в базах даних є досить складною задачею. Зазвичай, за критерій ефективності береться середня кількість порівнянь, необхідних для пошуку запису. На практиці це теоретичне середнє досить часто відрізняється від реальної середньої кількості порівнянь. Насамперед це пов'язано з тим, що ймовірності звертання до записів у файлах баз даних підпорядковані нерівномірним законам розподілу: пошук одних записів відбувається часто, інших дуже рідко. В роботі припускаємо, що можемо визначити закон розподілу, за яким розподілені записи. Це можна зробити, наприклад, маючи статистичну інформацію конкретної бази даних.

Проведемо дослідження ефективності методу m-паралельного послідовного перегляду [1] та двох варіантів методу m-паралельного блочного пошуку [2, 3] для різних законів розподілу ймовірностей звертання до записів, використовуючи коефіцієнти прискорення та паралельної ефективності. Серед законів розподілу ймовірностей розглянемо рівномірний і такі закони нерівномірного розподілу ймовірностей [4–6], як:

- "бінарний" розподіл

$$p_i = \frac{1}{2^i}, \quad i = \overline{1,\, N-1}, p_N = \frac{1}{2^{N-1}},$$

де p_i – ймовірність звертання до i-го запису, N – кількість записів у файлі;

- закон Зіпфа

$$p_i = \frac{1}{iH_N}, \quad i = \overline{1, N},$$

де $H_N = \sum_{k=1}^{N} \frac{1}{k}$;

- узагальнений закон розподілу

$$p_i = \frac{1}{i^c H_N^{(c)}}, \quad i = \overline{1, N},$$

де c $(0 < c < 1)$ – будь-який параметр і $H_N^{(c)} = \sum_{k=1}^{N} \frac{1}{k^c}$.

2. КОЕФІЦІЄНТИ ПРИСКОРЕННЯ ТА ПАРАЛЕЛЬНОЇ ЕФЕКТИВНОСТІ

Під час розробки паралельних алгоритмів принциповим моментом є аналіз ефективності використання процесорів. Цей аналіз зводиться до визначення оцінок отриманого прискорення процесу обчислень в результаті розпаралелення алгоритму (скорочення часу розв'язування задачі) та ефективності використання процесорів. Формування подібних оцінок проводиться стосовно вибраного обчислювального алгоритму. Аналіз можна дещо узагальнити, будуючи оцінки у випадку оптимальних алгоритмів розв'язування задач конкретного типу.

Коефіцієнтами ефективності паралельних алгоритмів є [7, 8]: прискорення (що показує у скільки разів швидше здійснюється розв'язування задачі при використанні декількох процесорів) і паралельна ефективність (що характеризує частку реального часу використання процесорів обчислювальної системи). Ці коефіцієнти широко використовуються в теорії і практиці паралельного програмування.

Прискорення (speedup), яке отримується при використанні паралельного алгоритму для m процесорів, у порівнянні з послідовним виконанням алгоритму визначається як відношення часу розв'язування задачі на однопроцесорній ЕОМ до часу виконання паралельного алгоритму на багатопроцесорній ЕОМ з m процесорами

$$S_m = T_1/T_m.$$

Паралельна ефективність (parallel efficiency) – це ефективність використання паралельним алгоритмом процесорів під час розв'язування задачі. Паралельна ефективність визначається співвідношенням

$$P_m = T_1/(mT_m) = S_m/m.$$

Величина паралельної ефективності визначає середню частку часу виконання алгоритму, протягом якої процесори реально задіяні для розв'язування задачі.

Можна показати, що в найкращому випадку $S_m = m$ і $P_m = 1$.

3. ДОСЛІДЖЕННЯ МЕТОДІВ ПАРАЛЕЛЬНОГО ПОШУКУ

Для методів паралельного пошуку у випадку їхніх оптимальних варіантів розрахуємо коефіцієнти прискорення та паралельної ефективності.

Коефіцієнти прискорення паралельного алгоритму пошуку обраховуватимемо за формулою

$$S_m = E_1/E_m.$$

де E_i – математичне сподівання кількості паралельних порівнянь, необхідних для пошуку запису у файлі, у випадку використання i процесорів. Обчисливши коефіцієнт прискорення, можемо побачити, у скільки разів швидше здійснюється розв'язування задачі у разі використання декількох процесорів.

Для визначення, наскільки ефективно використовуються паралельним алгоритмом процесори під час розв'язування задачі, необхідно визначити паралельну ефективність. Обчислимо коефіцієнт паралельної ефективності, використовуючи формулу

$$P_m = E_1/(mE_m) = S_m/m.$$

Нами обчислено коефіцієнти прискорення та паралельної ефективності методу m-паралельного послідовного перегляду та двох варіантів методу m-паралельного блочного пошуку у випадку оптимального розбиття на блоки. Результати обчислень зображені в наступних таблицях.

Для методу m-паралельного послідовного перегляду математичне сподівання кількості паралельних порівнянь у випадку різних законів розподілу ймовірностей звертання до записів має наступний вигляд [1]:

- рівномірний розподіл
$$E = \frac{1}{2}\left(\frac{N}{m} + 1\right);$$
- „бінарний" розподіл
$$E = \frac{2^m}{2^m - 1};$$
- закон Зіпфа
$$E = \frac{1}{H_N}\left(H_N + \frac{N}{m} - \frac{1}{2}\ln\frac{N}{m} - C_1\right),$$
де $C_1 = 0.5\ln 2\pi$;
- узагальнений розподіл
$$E = \frac{1}{H_N^{(c)}}\left[H_N^{(c)} + \frac{N^{1-c}}{1-c}\left(\frac{1-c}{2-c}n - \frac{\alpha^{(c)}(n)}{n^{1-c}}\right)\right],$$
де
$$\alpha^{(c)}(n) = H_n^{(c-1)} - \frac{1}{2-c}n^{2-c}.$$

В табл. 1, 2 наведені, відповідно, значення коефіцієнтів прискорення та паралельної ефективності для методу m-паралельного послідовного перегляду у випадку різної кількості процесорів, різних законів розподілу ймовірностей звертання до записів і $N = 10^6$.

Як бачимо з таблиць, отримано дуже хороші показники ефективності методу m-паралельного послідовного перегляду для всіх законів розподілу ймовірностей звертання до записів, крім „бінарного". Для m процесорів коефіцієнт прискорення S_m наближається до m, а коефіцієнт паралельної ефективності методу приблизно дорівнює одиниці. Звідси можемо зробити висновок, що ефективність алгоритму зростає пропорційно кількості процесорів і всі процесори

Таблиця 1.

Коефіцієнт прискорення методу m-паралельного послідовного перегляду для різних законів розподілу ймовірностей і різної кількості процесорів

m	Рівномірний	Узагальнений				Зіпфа	"Бінарний"
		c=0,2	c=0,4	c=0,6	c=0,8		
2	2,000	2,000	2,000	2,000	2,000	2,000	1,499
4	4,000	4,000	4,000	4,000	4,000	4,000	1,874
5	5,000	5,000	5,000	5,000	5,000	5,000	1,938
10	10,000	10,000	10,000	10,000	10,000	9,999	1,998
20	20,000	20,000	19,999	19,999	19,999	19,997	2,000
40	39,998	39,998	39,998	39,997	39,995	39,987	2,000
50	49,998	49,997	49,997	49,996	49,993	49,979	2,000
100	99,990	99,989	99,987	99,983	99,970	99,912	2,000

Таблиця 2.

Коефіцієнт паралельної ефективності методу m-паралельного послідовного перегляду для різних законів розподілу ймовірностей і різної кількості процесорів

m	Рівномірний	Узагальнений				Зіпфа	"Бінарний"
		c=0,2	c=0,4	c=0,6	c=0,8		
2	1,00000	1,00000	1,00000	1,00000	1,00000	0,99999	0,74963
4	1,00000	1,00000	1,00000	0,99999	0,99999	0,99998	0,46860
5	1,00000	1,00000	0,99999	0,99999	0,99999	0,99997	0,38760
10	0,99999	0,99999	0,99999	0,99998	0,99997	0,99993	0,19980
20	0,99998	0,99998	0,99997	0,99997	0,99994	0,99984	0,10000
40	0,99996	0,99996	0,99995	0,99993	0,99989	0,99967	0,05000
50	0,99995	0,99994	0,99993	0,99991	0,99985	0,99958	0,04000
100	0,99990	0,99989	0,99987	0,99983	0,99970	0,99912	0,02000

повністю задіяні впродовж виконання методу m-паралельного послідовного перегляду. Лише у випадку „бінарного" закону розподілу ймовірностей звертання до записів використання багатьох процесорів є недоцільним.

У випадку першого варіанту методу m-паралельного блочного пошуку математичне сподівання кількості порівнянь для розглянутих законів розподілу ймовірностей звертання до записів буде мати такий вигляд [2]:

- рівномірний розподіл

$$E = \frac{1}{2}(n+1) + \frac{1}{2}(s+1);$$

- „бінарний" розподіл

$$E = \frac{2^{ms} - s}{2^{ms} - 1} + \frac{2^m}{2^m - 1};$$

- закон Зіпфа

$$E = \frac{1}{H_N}(2H_N + n + (\frac{N}{mn} - 1)(\frac{1}{2}\ln n + C_1) - \frac{1}{2}\ln(\frac{N}{m}) - C_1)$$

де $C_1 = 0.5 \ln 2\pi$;

- узагальнений розподіл

$$E = \frac{1}{H_N^{(c)}}(2H_N^{(c)} - \frac{N^{1-c}}{1-c}(\frac{1-c}{2-c}n - (\frac{N}{mn} - 1)\frac{\alpha^{(c)}(n)}{n^{1-c}} + \frac{\alpha^{(c)}(N/m)}{(N/m)^{1-c}}))$$

де

$$\alpha^{(c)}(n) = H_n^{(c-1)} - \frac{1}{2-c}n^{2-c},$$

$$\alpha^{(c)}(ns) = H_{ns}^{(c-1)} - \frac{1}{2-c}(ns)^{2-c}.$$

Коефіцієнти паралельної ефективності для першого варіанту методу m-паралельного блочного пошуку у випадку різної кількості процесорів, різних законів розподілу ймовірностей звертання до записів і $N = 10^6$ наведено, відповідно, в табл. 3, 4.

Таблиця 3.

Коефіцієнт прискорення першого варіанту методу m-паралельного блочного пошуку для різних законів розподілу ймовірностей і різної кількості процесорів

m	Рівномірний	Узагальнений				Зіпфа	"Бінарний"
		c=0,2	c=0,4	c=0,6	c=0,8		
2	1,414	1,414	1,414	1,416	1,423	1,440	1,176
4	1,998	1,998	2,000	2,007	2,026	2,073	1,293
5	2,233	2,234	2,236	2,245	2,270	2,331	1,313
10	3,155	3,157	3,162	3,180	3,232	3,354	1,333
20	4,457	4,460	4,470	4,505	4,601	4,821	1,333
40	6,291	6,297	6,317	6,379	6,549	6,915	1,333
50	7,028	7,035	7,060	7,135	7,336	7,762	1,333
100	9,911	9,924	9,968	10,098	10,430	11,079	1,333

Проаналізувавши значення отриманого коефіцієнта прискорення для першого варіанту методу m-паралельного блочного пошуку, приходимо до висновку, що збільшення кількості процесорів веде лише до незначного збільшення ефективності. А значення коефіцієнта паралельної ефективності свідчить про незначну загруженість процесорів під час пошуку.

Явний вираз математичного сподівання для другого варіанту методу m-паралельного блочного пошуку у випадку різних законів розподілу ймовірностей звертання до записів має наступний вигляд [3]:

- рівномірний розподіл

$$E = \frac{1}{2}(n+1) + \frac{1}{2}(s+1);$$

Таблиця 4.

Коефіцієнт паралельної ефективності першого варіанту методу m-паралельного блочного пошуку для різних законів розподілу ймовірностей і різної кількості процесорів

m	Рівномірний	Узагальнений				Зіпфа	"Бінарний"
		c=0,2	c=0,4	c=0,6	c=0,8		
2	0,707	0,707	0,707	0,708	0,712	0,720	0,588
4	0,500	0,500	0,500	0,502	0,506	0,518	0,323
5	0,447	0,447	0,447	0,449	0,454	0,466	0,263
10	0,316	0,316	0,316	0,318	0,323	0,335	0,133
20	0,223	0,223	0,224	0,225	0,230	0,241	0,067
40	0,157	0,157	0,158	0,159	0,164	0,173	0,033
50	0,141	0,141	0,141	0,143	0,147	0,155	0,027
100	0,099	0,099	0,100	0,101	0,104	0,111	0,013

- „бінарний" розподіл

$$E = \frac{2^{m^2 s} - s}{2^{m^2 s} - 1} + \frac{2^m}{2^m - 1} - \frac{s}{2^{ms} - 1};$$

- закон Зіпфа

$$E = \frac{1}{H_N}\left[2H_N + n + \frac{1}{2}s\ln(nm) - \right.$$

$$\left. -\frac{1}{2}\ln n - \frac{1}{2}\ln(nms) + C_1(s-2)\right],$$

де $C_1 = 0.5\ln 2\pi$;

- узагальнений розподіл

$$E = \frac{1}{H_N^{(c)}}\left[2H_N^{(c)} + \frac{N^{1-c}}{1-c}\left(\frac{1-c}{2-c}n + \right.\right.$$

$$\left.\left. + \frac{s\alpha^{(c)}(nm)}{(nm)^{1-c}} - \frac{\alpha^{(c)}(n)}{n^{1-c}} - \frac{\alpha^{(c)}(nms)}{(nms)^{1-c}}\right)\right],$$

де

$$\alpha^{(c)}(nm) = H_n^{(c-1)} - \frac{1}{2-c}(nm)^{2-c},$$

$$\alpha^{(c)}(nms) = H_{ns}^{(c-1)} - \frac{1}{2-c}(nms)^{2-c}.$$

Коефіцієнти ефективності для другого варіанту методу m-паралельного блочного пошуку у випадку різної кількості процесорів та різних законів розподілу ймовірностей звертання до записів і $N = 10^6$ наведено, відповідно, в табл. 5, 6.

Одержані показники ефективності для другого варіанту методу m-паралельного блочного пошуку є значно кращими, ніж для першого варіанту. Як бачимо з табл. 5, ефективність другого варіанту зростає майже пропорційно збільшенню

Таблиця 5.

Коефіцієнт прискорення другого варіанту методу m-паралельного блочного пошуку для різних законів розподілу ймовірностей і різної кількості процесорів

m	Рівномірний	Узагальнений				Зіпфа	"Бінарний"
		c=0,2	c=0,4	c=0,6	c=0,8		
2	1,998	1,998	1,998	1,997	1,996	1,992	1,452
4	3,988	3,987	3,986	3,984	3,977	3,951	1,500
5	4,980	4,979	4,977	4,973	4,961	4,918	1,500
10	9,911	9,905	9,896	9,877	9,824	9,628	1,500
20	19,627	19,604	19,563	19,482	19,255	18,447	1,500
40	38,500	38,396	38,236	37,897	36,993	33,945	1,500
50	47,667	47,511	47,256	46,731	45,330	40,768	1,500
100	91,000	90,344	89,328	87,399	82,358	67,916	1,500

Таблиця 6.

Коефіцієнт паралельної ефективності другого варіанту методу m-паралельного блочного пошуку для різних законів розподілу ймовірностей і різної кількості процесорів

m	Рівномірний	Узагальнений				Зіпфа	"Бінарний"
		c=0,2	c=0,4	c=0,6	c=0,8		
2	0,9990	0,9989	0,9988	0,9986	0,9981	0,9960	0,7258
4	0,9970	0,9968	0,9965	0,9959	0,9942	0,9878	0,3750
5	0,9960	0,9958	0,9954	0,9945	0,9922	0,9836	0,3000
10	0,9911	0,9905	0,9896	0,9877	0,9824	0,9628	0,1500
20	0,9814	0,9802	0,9782	0,9741	0,9628	0,9224	0,0750
40	0,9625	0,9599	0,9559	0,9474	0,9248	0,8486	0,0375
50	0,9533	0,9502	0,9451	0,9346	0,9066	0,8154	0,0300
100	0,9100	0,9034	0,8933	0,8740	0,8236	0,6792	0,0150

кількості процесорів. А з табл. 6 слідує, що під час виконання паралельного алгоритму процесори загружені більше, ніж на 90%. Лише у випадку „бінарного" закону розподілу ймовірностей при збільшенні кількості процесорів ефективність паралельного алгоритму зменшується.

4. ВИСНОВКИ

Досліджено коефіцієнти прискорення та паралельної ефективності методу m-паралельного послідовного перегляду та двох варіантів методу m-паралельного блочного пошуку у випадку різних законів розподілу ймовірностей звертання до записів.

На основі одержаних показників ефективності приходимо до висновку, що ефективність методу m-паралельного послідовного перегляду та другого варіанту метода m-паралельного блочного пошуку зростає пропорційно кількості процесорів і всі процесори повністю задіяні впродовж виконання цих методів. Лише у випадку „бінарного" закону розподілу ймовірностей звертання до записів використання багатьох процесорів є недоцільним. У випадку першого варі-

анту методу m-паралельного блочного пошуку збільшення кількості процесорів веде лише до незначного збільшення ефективності, а коефіцієнт паралельної ефективності вказує про незначну загруженість процесорів під час здійснення пошуку.

1. *Лісовець В. Я. Цегелик Г. Г.* Метод m-паралельного послідовного перегляду записів та його використання для пошуку інформації у послідовних файлах баз даних // Фізико-математичне моделювання та інформаційні технології. – 2007. – Вип. 5. – С.109-119.
2. *Лісовець В. Я. Цегелик Г. Г.* Метод m-паралельного блочного пошуку записів у файлах баз даних та його ефективність // Відбір та обробка інформації. – 2007. – Вип. 27(103). – С. 87-92.
3. *Лісовець В. Я., Цегелик Г. Г.* Один з варіантів методу m-паралельного блочного пошуку записів і його ефективність //Фізико-математичне моделювання та інформаційні технології. – 2008. – Вип. 7. – С. 103-111.
4. *Кнут Д.* Искусство программирования для ЭВМ. Т.3: Сортировка и поиск. М.: Изд. дом. "Вильямс", 2000. 832 с.
5. *Мартин Дж.* Организация баз данных в вычислительных системах. М: Мир, 1980. 644 с.
6. *Цегелик Г.Г.* Организация и поиск информации в базах данных. Львов: Вища школа, 1987. 176 с.
7. *Воеводин В.В., Воеводин Вл.В.* Параллельные вычисления. СПб.: БХВ-Петербург, 2002. – 608 с.
8. *Alecu F.* Performance analysis of parallel algorithms. // Jornal of applied quantitative methods. – 2007. – Vol. 2, No. 1, Spring. – P. 129 – 134.

Одержано ..2010