

Paralelné algoritmy, časť č. 9

Optimálne paralelné triedenie

František Mráz

Kabinet software a výuky informatiky, MFF UK, Praha

Paralelné algoritmy, 2011/2012

1 Optimálne paralelné triedenie

Optimálne paralelné triedenie

Cole

paralelný mergesort pomocou *dobrych vzorkovacích postupností*

Definícia

- Prvok c je medzi prvkami a a $b \equiv_{df} a < c \leq b$.
- Nech e je prvok a L je množina. $\text{hodnosť}(e, L) =$ počet prvkov z L menších než e
- definujeme funkciu $R[A, B] : A \rightarrow \mathbb{N}$ tak, že $R[A, B] := \text{hodnosť}(e, B)$ – reprezentujeme poľom dĺžky $|A|$ tak, že $R[A, B][i] = \text{hodnosť}(A[i], B)$
- Utriedená postupnosť L sa nazýva dobrá vzorkovacia postupnosť (dvp) pre postupnosť $J \equiv_{df}$ medzi ľubovoľnými $k + 1$ po sebe idúcimi prvkami z $\{-\infty\} \cup L \cup \{\infty\}$ je maximálne $2k + 1$ prvkov z J
- Príklad: S utriedená postupnosť, S_1 je postupnosť obsahujúca každý druhý prvok z $S \Rightarrow S_1$ dpv S .

Mergewithhelp

function *mergewithhelp*(J, K, L)

{ L je dvp pre obidve utriedené postupnosti J a K }

begin

rozdeľ J na $J(1), J(2), \dots$, kde $J(i)$ je množina prvkov

ležiacich medzi $(i - 1)$ -vym a i -tym prvkom postupnosti L

analogicky K na $K(1), K(2), \dots$, kde \dots podľa L

for all $i, 1 \leq i \leq |L| + 1$ **in parallel do** $res_i := merge(J(i), K(i))$

mergewithhelp := $res_1 res_2 \dots res_{|L|+1}$

end

Vlastnosti dvp

Ozn. & označuje operáciu zlievania (merge)

Lemma

$$\left. \begin{array}{l} X \text{ dvp } X' \\ Y \text{ dvp } Y' \end{array} \right\} X, Y, X', Y' \text{ utriedené postupnosti.}$$

Potom

$$X \& Y \text{ dvp } X' \text{ a } X \& Y \text{ dvp } Y' .$$

- Pozor: $X \& Y$ nemusí byť dvp $X' \& Y'$!

$$\begin{array}{lll} X = [2, 7] & X' = [2, 5, 6, 7] & X \& Y = [1, 2, 7, 8] \\ Y = [1, 8] & Y' = [1, 3, 4, 8] & X' \& Y' = [1, 2, 3, 4, 5, 6, 7, 8] \end{array}$$

- $\text{redukcia}(X) =_{df}$ postupnosť obsahujúca každý štvrtý prvok z X

Vlastnosti dvp

Lemma

$\left. \begin{array}{l} X \text{ dvp } X' \\ Y \text{ dvp } Y' \end{array} \right\} X, Y, X', Y' \text{ utriedené postupnosti.}$

Potom medzi ľubovoľnými r po sebe idúcimi prvkami z $X \& Y$ je maximálne $2r + 2$ prvkov z $X' \& Y'$.

Dôkaz:

- môžeme predpokladať, že X, Y obsahujú $-\infty, +\infty$
- nech e_1, e_2, \dots, e_r je r po sebe idúcich čísel z $X \& Y - h$ prvkov z X a j prvkov z Y .
- BUNO $e_1 \in X$
- a) $e_r \in X$:
 - medzi e_1 a e_r je maximálne $2(h-1) + 1$ prvkov z X'
 - medzi e_1 a e_r je maximálne $2(j+1) + 1$ prvkov z Y' (e_1, \dots, e_r ležia medzi $j+2$ po sebe idúcimi prvkami z Y)
 - Celkovo: $2(h-1) + 1 + 2(j+1) + 1 = 2(h+j) + 2 = 2r + 2$

Vlastnosti dvp

Lemma

$$\left. \begin{array}{l} X \text{ dvp } X' \\ Y \text{ dvp } Y' \end{array} \right\} X, Y, X', Y' \text{ utriedené postupnosti.}$$

Potom medzi ľubovoľnými r po sebe idúcimi prvkami z $X \& Y$ je maximálne $2r + 2$ prvkov z $X' \& Y'$.

Dôkaz:

- môžeme predpokladať, že X, Y obsahujú $-\infty, +\infty$
- nech e_1, e_2, \dots, e_r je r po sebe idúcich čísel z $X \& Y - h$ prvkov z X a j prvkov z Y .
- BUNO $e_1 \in X$
- b) $e_r \in Y$:
 - medzi e_1 a e_r je maximálne $2h + 1$ prvkov z X'
 - medzi e_1 a e_r je maximálne $2j + 1$ prvkov z Y'
 - Gelkovo: $2r + 2$

Redukcia a zlievanie

Lemma

$$\left. \begin{array}{l} X \text{ dvp } X' \\ Y \text{ dvp } Y' \end{array} \right\} X, Y, X', Y' \text{ utriedené postupnosti.}$$

Potom medzi ľubovoľnými r po sebe idúcimi prvkami z $X \& Y$ je maximálne $2r + 2$ prvkov z $X' \& Y'$.

Použitie:

- Nech $Z = \text{redukcia}(X \& Y)$ a $Z' = \text{redukcia}(X' \& Y')$
- e_1, \dots, e_{k+1} po sebe idúce prvky zo Z
 medzi nimi je maximálne $4k + 1$ prvkov $X \& Y$ (vrátane e_1 a e_{k+1})
 medzi týmito $4k + 1$ prvkami leží maximálne $8k + 4$ prvkov z $X' \& Y'$, teda po redukcii maximálne $(8k + 4) \cdot \frac{1}{4} = 2k + 1$ prvkov zo Z'
 $\Rightarrow Z \text{ dvp } Z'$.

Schéma algoritmu triedenia

- predpokladáme počet vst. prvkov rovný mocnine 2
binárny strom – uzly budú obsahovať utriedené postupnosti prvkov
 - **Inicializácia:** i -ty list obsahuje i -ty vstup, vnútorné uzly obsahujú prázdne zoznamy
 - **Výpočet:**
 val_v postupnosť uložená v uzle v
 T_v podstrom s koreňom v
 $list_v$ zoznam prvkov v listoch stromu T_v
 - úloha uzlu v : utriediť postupnosť $list_v$
 val_v bude vždy podpostupnosť postupnosti $list_v$
- vrchol v je $\begin{cases} \text{úplný} & \text{ak } |val_v| = |list_v| \\ \text{neúplný} & \text{inak} \end{cases}$

Priebeh výpočtu v uzle v

- v dostáva dva prúdy utriedených postupností
 X_1, X_2, \dots, X_r z ľavého syna
 Y_1, Y_2, \dots, Y_r z ľavého syna
 vždy X_i a Y_i zároveň.
- následne počíta prúd utriedených postupností

$$(val_v)_i := \text{mergewithhelp}(X_i, Y_i, (val_v)_{i-1})$$

a vysiela výstupný prúd utriedených postupností Z_1, \dots, Z_{r+1}

- vo všetkých prúdoch je dĺžka nasledujúcej postupnosti dvojnásobná oproti predchádzajúcej
- ak je v neúplný, tak otcovi posíla *redukcia*(val_v)
 ak je v úplný, tak pošle najprv každý druhý prvok a potom všetky prvky a prestane pracovať.

Počet procesorov

- procesory “putujú” spolu s triedenými prvkami
- nech vrcholy do výšky v_1 dopočítali
- okrem úplnej úrovne je vždy v nasledujúcej vyššej vrstve v uzle postupnosť maximálne polovičnej dĺžky než v nižšej vrstve
- počet vrcholov vo výške h je $2^{\log n - h}$, dĺžka $|val_v|$ vo výške h je $2^{2v_1 - h}$
- počet procesorov v aktívnych vrcholoch stromu je maximálne

$$\begin{aligned} \sum_{h=v_1}^{2v_1} 2^{\log n - h} \cdot 2^{2v_1 - h} &\leq 2^{\log n + 2v_1} \sum_{h=v_1}^{2v_1} 2^{-2h} = n 2^{2v_1} \sum_{h=0}^{v_1} 2^{-2(h+v_1)} = \\ &= n 2^{2v_1 - 2v_1} \sum_{h=0}^{v_1} 2^{-2h} \leq n \frac{1}{1 - \frac{1}{4}} = n \frac{4}{3} \end{aligned}$$

plus úplná vrstva = $O(n)$

- celkovo čas $O(\log n)$ s n procesormi

Invarianty

- **Hlavný invariant:**

- $\forall i$ ak X_i dvp X_{i+1} a Y_i dvp Y_{i+1} potom Z_i dvp Z_{i+1} ;

- $|X_i| = |Y_i| = |Z_i| = \frac{|X_{i+1}|}{2} = \frac{|Y_{i+1}|}{2} = \frac{|Z_{i+1}|}{2}$

- indukciou

špeciálne sa ošetrí Z_r a Z_{r+1} (nejde pomocou lemy o redukciách), ale

- Z_{r-1} je každý druhý zo Z_r , teda Z_{r-1} dvp Z_r

- Z_r je každý druhý zo Z_{r+1} , teda Z_r dvp Z_{r+1}

- ak X ukončí prúd v čase t , tak jeho otec skončí v čase $t + 3 \Rightarrow$ celkový počet štádií je $3 \log n$

- **Pomocný invariant:**

- Ak $[S_1, \dots, S_p]$ je prúd vstupujúci alebo vystupujúci z nejakého uzlu, tak poznáme aj prúd krížových odkazov $R[S_{i+1}, S_i]$

- Pozn.: pre každú postupnosť S poznáme $R[S, S]$

Hodnosť prvku v postupnosti dĺžky k pomocou k procesorov

Lemma

Nech $S = [b_1, \dots, b_k]$ je utriedená postupnosť. Potom hodnosť ľubovoľného prvku a v S sa dá spočítať v konštantnom čase pomocou $O(k)$ procesorov na CREW PRAMe.

Dôkaz: nech $b_0 = -\infty$, $b_{k+1} = +\infty$
 for all $0 \leq i \leq k$ in parallel do
 if $b_i < a \leq b_{i+1}$ then hodnosť(a, S) := i

Lemma

Nech S_1, S_2 a $S = S_1 \& S_2$ sú utriedené postupnosti, $S_1 \cap S_2 = \emptyset$. Potom $R[S_1, S_2]$ a $R[S_2, S_1]$ sa dá spočítať v čase $O(1)$ pomocou $O(|S|)$ procesorov na CREW PRAM.

Dôkaz: poznáme $R[S, S] \Rightarrow$
 $\text{hodnosť}(a, S_2) = \text{hodnosť}(a, S_1 \& S_2) - \text{hodnosť}(a, S_1)$

Hlavná technická lemma I

Lemma

Nech sú dané utriedené postupnosti

$X, Y, U = X \& Y, X', Y', X \text{ dvp } X', Y \text{ dvp } Y'$. Ak sú dané krížové odkazy $R[X', X], R[Y', Y]$, tak môžeme spočítať krížové odkazy $R[X', U], R[Y', U]$ v čase $O(1)$ pomocou $O(|X| + |Y|)$ procesorov na CREW PRAM.

Dôkaz: Výpočet $R[X', X \& Y]$ (výpočet $R[Y', X \& Y]$ analogicky), nech $X = [a_1, \dots, a_k]$

a) X' rozdelíme na segmenty

$X'(1) \dots$ prvky ležiace medzi $-\infty, a_1$

$X'(2) \dots$ prvky ležiace medzi a_1, a_2

\dots

čas $O(1)$ s $O(|X|)$ procesorov (poznáme $R[X', X]$)

Hlavná technická lemma I

Lemma

Nech sú dané utriedené postupnosti

$X, Y, U = X \& Y, X', Y', X \text{ dvp } X', Y \text{ dvp } Y'$. Ak sú dané krížové odkazy $R[X', X], R[Y', Y]$, tak môžeme spočítať krížové odkazy $R[X', U], R[Y', U]$ v čase $O(1)$ pomocou $O(|X| + |Y|)$ procesorov na CREW PRAM.

Dôkaz: Výpočet $R[X', X \& Y]$

b) $U(i)$ prvky z Y , ktoré ležia medzi a_{i-1} a a_i v U

$\forall x \in X'$ spočítame hodnotu (x, U) :

for all $i, 1 \leq i \leq k + 1$ in parallel do

for all $x \in X'(i)$ do

begin spočítame hodnotu $(x, U(i))$ (prvá lemma o 2 slajdy dozadu)

$\text{hodnota}(x, U) := j + \text{hodnota}(x, U(i))$

j je poradie a_i v U

end

Pre každé i použijeme $|U(i)|$ procesorov, $|X'(i)| \leq 3$ pretože

$X \text{ dvp } X'$

\Rightarrow čas $O(1)$

Hlavná technická lemma II

Lemma

Nech sú dané utriedené postupnosti

$X, Y, U = X \& Y, X', Y', X \text{ dvp } X', Y \text{ dvp } Y'$. Ak sú dané krížové odkazy $R[X', X], R[Y', Y]$, tak môžeme spočítať krížové odkazy $R[U, X'], R[U, Y']$ v čase $O(1)$ pomocou $O(|X| + |Y|)$ procesorov na CREW PRAM.

Dôkaz: Výpočet $R[X \& Y, X']$ ($R[X \& Y, Y']$ analogicky)

- najprv spočítame $R[X, X']$

$R[X, X'](j) := \text{index max. prvku z } X'(j):$

- o každom prvku z $X \& Y$ viem či pochádza z X alebo z Y
- ak $a_j \in X_i$: $R[X \& Y, X'](a_j)$ spočítame z $R[X, X'](a_j)$
- ak $a_j \in Y_i$: $R[X \& Y, X'](a_j)$ spočítame z hodnot'(a_j, X) a $R[X, X']$
 $\text{hodnost'}(a_j, X) := \text{hodnost'}(a_j, X \& Y) - \text{hodnost'}(a_j, Y)$