ساختمان داده ها

Deterministic Computation of Order Statistic

محاسبه قطعی آماره ترتیبی

مدرس: غیاثی‌شیرازی

دانشگاه فردوسی مشهد

Randomized Selection

$T(n)$ زمان اجرا (حداکثر) برای آرایه به طول n

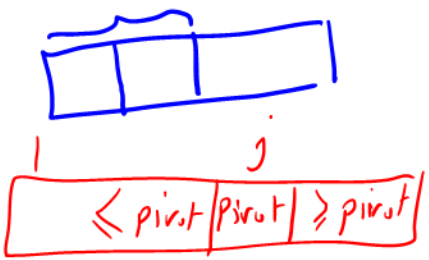
$T(n) \in \Omega(n)$

و

$T(n) \in O(n)$

$\Updownarrow$

$T(n) \in \Theta(n)$



~~R~~ D Select(A, n, i)

انتخاب محور با الگوریتم افراز به دسته‌های ۵ تایی

if n == 1

  return A[1]

pivot = Choose Pivot(A, n) $\Theta(n) + T(n/5)$ $\leq cn$

j = Partition(A, n, pivot) $\Theta(n)$

if i == j

  return A[j]

elif i < j

  ~~R~~ D Select(A[بخش اول], j-1, i)

else

  ~~R~~ D Select(A[بخش دوم], n-j, i-j)

حداکثر

$T(7n/10)$

$T(n) \leq cn + T(n/5) + T(7n/10)$

$T(1) = 1$

$$T(n) \leq cn + T(n/5) + T(7n/10)$$

$$T(1) = 1$$

گزاره: $\exists a > 0 \;\; \exists n_0 \in \mathbb{N} \;\; \forall n > n_0 \;\; T(n) \leq an$ یعنی $T(n) \in O(n)$

$a = 10c$ ، $n_0 = 1$

اثبات با استقرای قوی

فرض استقرا: برای هر k که $k < n$ داریم: $T(k) \leq ak$

ثابت می‌کنیم $T(n) \leq an$

$$T(n) \leq cn + T(n/5) + T(7n/10)$$

$$\leq cn + \frac{an}{5} + \frac{7}{10}an = cn + \frac{9}{10}an \leq an$$

باید نشان دهیم

پس باید $10c \leq a$ باشد.

$$\leq \frac{10c + 90c}{10} n = 10cn = an$$

$\Leftarrow \; T(n) \leq an$

$$n + (n-1) + (n-2) + \cdots + 1 \in \Theta(n^2)$$

---

k عدد ثابت

$$n + (n-k) + (n-2k) + \cdots$$

$$\sum_{i=1}^{\lfloor n/k \rfloor} (n - ik) = \sum_{i=1}^{\lfloor n/k \rfloor} n - k \sum_{i=1}^{\lfloor n/k \rfloor} i = n \lfloor n/k \rfloor - \cancel{k} \frac{\lfloor n/\cancel{k} \rfloor (\lfloor n/k \rfloor + 1)}{2}$$

$$\approx \frac{n^2}{k} - \frac{n^2}{2k} = \frac{n^2}{2k} \in \Theta(n^2)$$

$0 < \alpha < 1$ $\in \Theta(n)$

$$n + \alpha n + \alpha^2 n + \cdots = n(1 + \alpha + \alpha^2 + \cdots) = \frac{n}{1-\alpha}$$

Deterministic Selection انتخاب محور

۱. داده‌ها را به دسته‌های 5 تایی تقسیم می‌کنیم

۲. داده‌های هر دسته را مرتب می‌کنیم

۳. عناصر میانه هر دسته 5 تایی را در آرایه B قرار می‌دهیم

۴. Pivot $\leftarrow$ DSelect$\left(B, \lceil n/5 \rceil, \lceil n/10 \rceil\right)$

A

3	40	5	25	19	5	6	17	15
2	30	60	9	39	16	8	33	13
4	20	3	13	64	29	5	26	50
5	10	17	42	7	14	85	4	15
1	50	2	6	18	27	64	13	8

پس از مرتب سازی
دسته‌های 5 تایی

B

5	50	60	42	64	29	85	33	50
4	40	17	25	39	27	64	26	15
3	30	5	13	19	16	8	17	15
2	20	3	9	18	14	6	13	13
1	10	2	6	7	5	5	4	8

$\leftarrow$ DSelect(B, 9, 5)

3, 5, 8, 13, $\boxed{15}$, 16, 17, 19, 30

پس از مرتب سازی دسته‌های 5 تایی

5	50	60	42	64	29	85	33	50
4	40	17	25	39	27	64	26	15
3	30	5	13	19	16	8	17	15
2	20	3	9	18	14	6	13	13
1	10	2	6	7	5	5	4	8

B

5	60	85	42	50	29	33	64	50
4	17	64	25	15	27	26	39	40
3	5	8	13	15	16	17	19	30
2	3	6	9	13	14	13	18	20
1	2	5	6	8	5	4	7	10

$\geq$ pivot

≤ 15

در ادامه حداکثر طول آرایه $7/10\,n$ است

مرغ و تخم مرغ

- برای آنکه الگوریتم محاسبه i امین آماره ترتیبی به خوبی عمل کند باید محور را به نحو مناسبی انتخاب کنیم که آرایه به دو بخش مناسب شکسته شود.
- انتخاب بهینه محور این است که مقدار آن برابر میانه باشد که $\left\lfloor \frac{n+1}{2} \right\rfloor$ امین آماره ترتیبی است.
- ایده:
  - محور را به نحو مناسبی (نه الزاما میانه) با استفاده از همان الگوریتم محاسبه i امین آماره ترتیبی به دست آوریم.

الگوریتم انتخاب محور

- آرایه A را به صورت n/5 آرایه ۵ تایی در نظر بگیر و هر آرایه را مرتب کن.
- آرایه C را با مقادیر میانه آرایه های ۵ تایی پر کن.
- میانه C را با استفاده از الگوریتم پیدا کردن i امین آماره ترتیبی محاسبه کن.
- مقدار میانه C را به عنوان محور انتخاب کن

الگوریتم محاسبه i امین آماره

- DSelect(array A, length n, order statistic i)

1. p = ChoosePivot(A,n);
2. Partition A around p
3. If j = i return p
4. If j < i return DSelect(1st part of A, j-1, i)
5. [else if j > i] return DSelect(2nd part of A, n-j, i-j)

الگوریتم قطعی انتخاب محور

- ChoosePivot(array A, length n, order statistic i)

1. Break A into groups of 5, sort each group
2. C = the n/5 “middle elements”
3. p = DSelect(C,n/5,n/10)
4. return p

الگوریتم DSelect بدون فراخوانی تابع ChoosePivot

- DSelect(array A, length n, order statistic i)

1. Break A into groups of 5, sort each group
2. C = the n/5 “middle elements”
3. p = DSelect(C,n/5,n/10) [computes median of C]
4. Partition A around p
5. If j = i return p
6. If j < i return DSelect(1st part of A, j-1, i)
7. [else if j > i] return DSelect(2nd part of A, n-j, i-j)

زمان اجرای الگوریتم DSelect

- **قضیه**: برای هر آرایه ورودی با اندازه n زمان اجرای الگوریتم Dselect از مرتبه O(n) است.
- اخطار: در عمل الگوریتم Rselect بهتر است، زیرا
  1. نیاز به حافظه اضافی ندارد.
  2. ضریب ثابت مرتبه زمانی آن کمتر است.

تحلیل زمان اجرای الگوریتم DSelect

- DSelect(array A, length n, order statistic i)

Θ(n)	1. Break A into groups of 5, sort each group
Θ(n)	2. C = the n/5 “middle elements”
T(n/5)	3. p = DSelect(C,n/5,n/10) [computes median of C]
Θ(n)	4. Partition A around p
Θ(1)	5. If j = i return p
T(?)	6. If j < i return DSelect(1st part of A, j-1, i)
	7. [else if j > i] return DSelect(2nd part of A, n-j, i-j)

$$T(n) \leq cn + T\left(\frac{n}{5}\right) + T(?\,)$$

لم کلیدی

- طول آرایه در فراخوانی بازگشتی دوم در خطوط ۶ یا ۷ حداکثر برابر $\frac{7}{10}n$ است.

اثبات لم کلیدی

- فرض کنید داده های ۵ تایی مرتب شده در ستون های یک جدول قرار گرفته اند و این ستون ها بر اساس مقدار میانه (ردیف سوم) مرتب شده اند.

۱۱۰۰		**۴۰۰**	**۳۸۰**	**۲۳۰**	**۹۸**	**۸۰۰**
۱۰۰۰		۱۷۰	۱۴۵	۱۲۸	۷۲	۱۹۰
۸۰۰	...	۱۲۰	۷۶	۴۳	۳۵	۱۶
۸		۹۰	۶۹	۹	۱۲	۱۵
۲		۶۰	۴	۵	۱	۷

اثبات لم کلیدی

- عنصر محور از تمام عناصر سمت چپ و پایین بزرگ تر است.
- عنصر محور از تمام عناصر سمت راست و بالا کوچک تر است.

۸۰۰	**۹۸**	**۲۳۰**	**۳۸۰**	**۴۰۰**		**۱۱۰۰**
۱۹۰	۷۲	۱۲۸	۱۴۵	۱۷۰		۱۰۰۰
۱۶	۳۵	۴۳	۷۶	۱۲۰	...	۸۰۰
۱۵	۱۲	۹	۶۹	۹۰		۸
۷	۱	۵	۴	۶۰		۲

اثبات لم کلیدی

- فرض کنیم آرایه به $k \approx \frac{n}{5}$ دسته ۵ تایی تقسیم شده است.
- عنصر محور در ستون $\frac{k}{2}$ قرار دارد و مقدار آن تقریبا از 3/5 عناصر سمت چپ جدول بزرگ تر یا مساوی است. همچنین مقدار آن تقریبا از 3/5 عناصر سمت راست جدول کوچک تر یا مساوی است.

۸۰۰	**۹۸**	**۲۳۰**	**۳۸۰**	**۴۰۰**		**۱۱۰۰**
۱۹۰	۷۲	۱۲۸	۱۴۵	۱۷۰		۱۰۰۰
۱۶	۳۵	۴۳	۷۶	۱۲۰	...	۸۰۰
۱۵	۱۲	۹	۶۹	۹۰		۸
۷	۱	۵	۴	۶۰		۲

اثبات لم کلیدی

- بنابراین با توجه به اینکه عناصر سمت چپ آرایه تقریبا نصف عناصر کل آرایه را تشکیل می دهند، عنصر محور از 0.3 عناصر کل آرایه بزرگ تر است.
- به طور مشابه عنصر محور از 0.3 عناصر کل آرایه کوچک‌تر است.
- بنابراین طول آرایه مورد بررسی در فراخوانی تابع DSelect در خط ۶ یا ۷ حداکثر برابر $\frac{7}{10}n$ می باشد.

تکمیل اثبات خطی بودن زمان اجرای DSelect

- طبق لم کلیدی، زمان اجرای الگوریتم در رابطه بازگشتی زیر صدق می کند:

$$T(1) = 1$$

$$T(n) \leq cn + T\left(\frac{n}{5}\right) + T\left(\frac{7}{10}n\right)$$

- با استقرا نشان می دهیم که $T(n) \leq an$
- که a مقداری ثابت است.

اثبات استقرایی

- ابتدای استقرا:

$$T(1) = 1 \leq a.1$$

- فرض استقرا: برای هر k<n داریم:

$$T(k) \leq a.k$$

- حکم استقرا: باید نشان دهیم:

$$T(n) \leq a.n$$

اثبات استقرایی

- **اثبات**: طبق تحلیل الگوریتم داریم:

$$T(k) \leq cn + T\left(\frac{n}{5}\right) + T\left(\frac{7}{10}n\right)$$

$$\leq cn + \frac{an}{5} + \frac{7an}{10} = \frac{(10c + 2a + 7a)n}{10}$$

- کافی است a را برابر 10c بگیریم تا داشته باشیم:

$$T(n) \leq \frac{10a}{10}n = a.n$$