מבני נתונים 17 ~ בעיית הבחירה

שחר פרץ

11 ביוני 2025

הבעיית הבחירה: בהינתן n איברים, עם סדר מלא עליהם, כיצד נוכל למצוא את המינימום ה־k. עבור עץ ממויין, כבר ראינו איך פותרים אותה.

עד לפני 50 שנה עוד התעסקו עם הבעיה הזו. נקרא ל"Dyanmic setting" מצב בו כל פעם שמקבלים את הבעיה צריך לבנות משהו ואז יש עלות תחזוק. בניגוד, פתרון סטטי יהיה למדיין רשימה פעם יחידה בלי בעיה, ומשם ואילך פתרון ב־$O(1)$.

נוכל לנסות לפתור באמצעות heaps, בתרכיל הבית לדוגמה נוכיח שאפשר לפתור את זה כך ב־ב־$O(n + k\log k)$.

QUICKSELECT . . . . (1)

1.1 Randomized QuickSort

נציג אלגוריתם quicksort של $O(n)$ בתוחלת ואף $O(n)$ W.C. שימו לב להבדל בקורס הזה בין אלגוריתם בתוחלת ואלגוריתם בממוצע.

אלג' quickselect יבחר k אקראי ויסדר לפי מי שגדול ומי שקטן ממנו, ואז יפעיל את האלגו' על ה־partition הרלוונטי (כתלות בכמה איברים היו לפניו). W.C. $O(n^2)$, B.C> $O(n)$.

בהקשר הבא, n_i גודל המערך בחלוקה ה־i, הוא המ"מ שלנו:

$$\mathbb{E}[x] = \mathbb{E}\left[\sum_{i=1}^{n} n_i\right] = \sum_{i=1}^{n}\mathbb{E}[n_i] \overset{!}{\leq} \sum_{i=1}^{\infty} nc^i = O(n) \quad (c<1)$$

לעמשה, מה שצריך להראות הוא ש־$\mathbb{E}[n_{i+1}] \leq c\cdot\mathbb{E}[n_i]$, כלומר התוחלת של $i+1$ קטנה מזו של i בקבוע.

ננסה לנתח כמה מקרים פשוטים כדי לחסום את $\mathbb{E}[n+i]$.

- עבור $k=1$, נבחין ש־

$$\mathbb{E}[n_1] = \sum n_1(\text{מאורע})\cdot\underbrace{P(\text{מאורע})}_{\frac{1}{n}} = \frac{\sum_{i=0}^{n} i}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2} = O(n)$$

- עבור המקרה הכללי: נפלג למקרים:
  - במקרה ש־$P \leq z_k$, כאשר z_k המיקום של k במערך. אזי p יכול להיות כל אחד מ־$z_1 \dots z_k$ כאשר הסבירות לכולם שווה:

$$\mathbb{E} = \sum n_1(\text{מאורע})P(\text{מאורע}) \geq \frac{k}{2}$$

  כי בממוצע מס' האלמנטים שיוסרו בין 1 ל־k הוא $\frac{k}{2}$.

  - במקרה ש־$p > z_k$, נקבל בממוצע $\frac{n-k}{2} \geq \frac{(n-k)+1}{2}$.

  ניעזר בנוסחת ההסתברות המותנה בעבור תוחלות:

$$P(A) = \sum_{i=1}^{n} P(A \mid F_i)\cdot P(F_i) \qquad \mathbb{E}[x] = \sum_{i=1}^{n}\mathbb{E}[X \mid F_i]\cdot P(F_i)$$

  נוכל להשתמש בחסימה הבאה:

$$n_{i+1} \leq n_i - \#\text{removed} \overset{!}{\leq} n_i$$

$$\begin{aligned}\mathbb{E}[\#\text{removed}] &= E[\#\text{removed} \mid p \leq z_k]P(p\leq z_k) + \mathbb{E}[\#\text{removed} \mid p > z_k]\cdot P(p > z_k)\\ &= \frac{k}{2}\cdot\frac{k}{n} + \frac{n-k}{2}\cdot\frac{n-k}{n} = \frac{1}{2n}(k^2+(n-k)^2) \overset{\forall k}{\geq} \frac{n}{4}\end{aligned}$$

הא"ש ניתן להוכיח באמצעות גזירה לפי k, כמו שעשינו בקורס B. סה"כ הראינו ש־$n_{i+1} \leq n_i - \frac{1}{n}n_i = \frac{3}{4}n_i$ כדרוש. סה"כ:

$$\mathbb{E}[x] = \mathbb{E}\left[\sum_{i=0}^{n} n_i\right] = \sum_{i=0}^{n} e_i \leq n\left(\sum_{i=1}^{\infty}\left(\frac{3}{4}\right)^i\right) = 4n = O(n)$$

1.2 Deterministic QuickSelect (MedoMed)

ראינו בעבר ש־$T(\alpha n) + T(\beta n) + O(n) = T(n) = O(n)$ עבור $\alpha + \beta < 1$. אז נצטרך לבחור משהו שהוא "באיזור של" החציון. נעשה את זה ע"י חלוקת המערך לחמישיות ב־$O(n)$, ואז עבור האיבר ה־i ברשימות האלו נמצא את "החציון של החציונים" ב־$T(\frac{n}{5})$. ואז יש עוד דברים שלא הבנתי ואמור לצאת $T(n) = c_1 n + T(n/5) + c_2(n) + T(0.7n)$.

המשמעות של זה היא שאפשר לעשות quicksort ב־W.C. של $n \log n$ באמצעות מציאת median.

....

שחר פרץ, 2025