
Datenstrukturen, Algorithmen und Programmierung 2

Amin Coja-Oghlan

June 10, 2022

Lehrstuhl Informatik 2
Fakultät für Informatik

Mediane und das Auswahlproblem

Das Auswahlproblem

- gegeben ist ein Array $A = (A_1, \dots, A_n)$ und eine Zahl $1 \leq \ell \leq n$
- die Elemente sind vergleichbar
- das Ziel ist, das ℓ -te Element von A in aufsteigender Reihenfolge zu finden

Mediane und das Auswahlproblem

Lösung via Sortieren

- offenbar könnten wir A in Zeit $O(n \log n)$ sortieren und einfach den ℓ -ten Eintrag des sortierten Arrays ausgeben
- aber für $\ell = 1$ und $\ell = n$ kann das Problem in Zeit $O(n)$ gelöst werden
- geht das auch für allgemeine ℓ ?

Mediane und das Auswahlproblem

Definition

Ein **Median** eines Arrays $A = (A_1, \dots, A_n)$ ist ein Element A_i von A , so daß

$$|\{i \in [n] : A_i < m\}| \leq \frac{n}{2} \quad \text{und} \quad |\{i \in [n] : A_i > m\}| \leq \frac{n}{2}.$$

- wenn wir das Array sortieren, ist das Element in Position $\lceil (n+1)/2 \rceil$ ein Median
- wenn n ungerade ist und alle Elemente verschieden sind, ist der Median eindeutig
- wenn n gerade ist, gibt es mindestens zwei Mediane

Mediane und das Auswahlproblem

Select(**A**, ℓ)

1. Falls $n = 1$, gib A_1 aus.
2. Unterteile **A** in $k = \lfloor n/5 \rfloor$ Teilarrays $T_1, \dots, T_{\lfloor n/5 \rfloor}$ zu je 5 Elementen und, falls n nicht durch 5 teilbar ist, ein weiteres Teilarray $T_{\lfloor n/5 \rfloor + 1}$ auf.
3. Setze $N = \lfloor n/5 \rfloor$, falls n durch 5 teilbar ist, und $N = \lfloor n/5 \rfloor + 1$ sonst.
4. Finde in jedem dieser Teilarray T_i einen Median m_i .

Mediane und das Auswahlproblem

Select(\mathbf{A}, ℓ)

5. Wende Select rekursiv an, um einen Median m von $\mathbf{m} = (m_1, \dots, m_N)$ zu finden.
6. Bestimme $\mathbf{K} = (K_1, \dots, K_{n'})$, $\mathbf{M} = (M_1, \dots, M_{n''})$, $\mathbf{G} = (G_1, \dots, G_{n'''})$, die die Elemente von \mathbf{A} kleiner/gleich/größer m enthalten.
7. Falls $n' \geq \ell$, führe Select(\mathbf{K}, ℓ) aus;
8. sonst, falls $n' + n'' \geq \ell$, gib m aus;
9. sonst führe Select($\mathbf{G}, \ell - n' - n''$) aus.

Mediane und das Auswahlproblem

Satz

$\text{Select}(\mathbf{A}, \ell)$ hat Laufzeit $O(n)$.

Mediane und das Auswahlproblem

Zusammenfassung

- Select bestimmt das ℓ -te Element in linearer Zeit
- es gibt einen einfacheren randomisierten Algorithmus für dieses Problem
- dieser ähnelt Quicksort, aber “verfolgt nur einen Zweig”