Decision Trees, Adversary Argument, and Amortized Analysis

Hengfeng Wei

hfwei@nju.edu.cn

April 11 $\sim$ April 12, 2017

Decision Trees, Adversary Argument, and Amortized Analysis

Algorithm analysis

- Given a problem P
- Design algorithms $A, A', \ldots$
- Input space $\mathcal{X}_n$: inputs of size n

$$T_A(n) \triangleq \max_{X \in \mathcal{X}_n} T_A(X)$$

$$T_P(n) \triangleq \min_{A \text{ solves } P} T_A(n) = \min_{A \text{ solves } P} \max_{X \in \mathcal{X}_n} T_A(X)$$

Algorithm analysis

Sorting:

$$
\begin{aligned}
& n! \\
& \Longrightarrow n^2 \\
& \Longrightarrow n \log n \\
n \log n \Longleftarrow & \\
& n
\end{aligned}
$$

Selection (median):

$$
\begin{aligned}
& n^2 \\
& \Longrightarrow n \log n \\
& \Longrightarrow 16n \\
& \Longrightarrow 2.95n \\
2n \Longleftarrow & \\
\frac{3n}{2} - \frac{3}{2} \log n \Longleftarrow & \\
& n
\end{aligned}
$$

Decision Trees, Adversary Argument, and Amortized Analysis

K-sorted (Problem 2.9)

$$1,\ 2,\ 4,\ 3; \quad 7,\ 6,\ 8,\ 5; \quad 10,\ 11,\ 9,\ 12; \quad 15,\ 13,\ 16,\ 14$$

1-sorted?
2-sorted?
n-sorted?

1-sorted $\to$ 2-sorted $\to$ 4-sorted $\to \cdots \to$ n-sorted

Quicksort (with median) stops after the $\log k$ recursions.

$$O(n \log k)$$

K-sorted (Problem 2.9)

$$\Omega(n \log k)$$

$$L = \binom{n}{n/k, \ldots, n/k} = \frac{n!}{\left((\frac{n}{k})!\right)^k}$$

$$H \geq \log\left(\frac{n!}{\left((\frac{n}{k})!\right)^k}\right)$$

K-sorted (Problem 2.9)

Sorting the k-sorted array.

$$O(n \log \frac{n}{k})$$

$$L = ((\frac{n}{k})!)^k$$

$$H \geq \log((\frac{n}{k})!)^k = \Omega(n \log \frac{n}{k})$$

Decision Trees, Adversary Argument, and Amortized Analysis

Horse Racing

Example (Horse Racing)

- 25 horses
- Round: ≤ 5 horses race
- Goal: Find $\#1, \#2, \#3$ fastest.

$$8 \implies 7 \quad (a_2, a_3, b_1, b_2, c_1)$$

$$(< 5) \implies (= 5) \quad (\#1) \implies (= 6) \quad (\#2)$$

Finding patterns in bit strings

Example (Finding patterns in bit strings)

- Bit string $A[1 \ldots n]$
- Bit pattern 01
- Question: checking every bit?

n is odd: checking $A[2, 4, \ldots, n-1]$

n is even: adversary argument

Decision Trees, Adversary Argument, and Amortized Analysis

Amortized analysis

Amortized analysis is
an algorithm analysis technique for
analyzing a sequence of operations
irrespective of the input to show that
the average cost per operation is small, even though
a single operation within the sequence might be expensive.

Methods for amortized analysis: the summation method

$$o_1, o_2, \ldots, o_n$$

$$c_1, c_2, \ldots, c_n$$

$$(\sum_{i=1}^{n} c_i)/n$$

Summation method: array doubling revisited

On any sequence of n INSERT ops on an initially empty array.

$$
\begin{array}{ccccccccccc}
o_i: & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
c_i: & 1 & 2 & 3 & 1 & 5 & 1 & 1 & 1 & 8 & 1
\end{array}
$$

$$
c_i = \begin{cases} (i-1)+1 = i & \text{if } i-1 \text{ is an exact power of 2} \\ 1 & \text{o.w.} \end{cases}
$$

$$
\sum_{i=1}^{n} c_i \le n + \sum_{j=0}^{\lceil \lg n \rceil - 1} 2^j = n + (2^{\lceil \lg n \rceil} - 1) \le n + 2n = 3n
$$

$$
\forall i, \hat{c}_i = 3
$$

Methods for amortized analysis: the accounting method

$$o_1, o_2, \ldots, o_n$$

$$c_1, c_2, \ldots, c_n$$

$$a_1, a_2, \ldots, a_n$$

$$\hat{c_i} = c_i + a_i, a_i >=< 0.$$

$$\forall n, \sum_{i=1}^{n} c_i \le \sum_{i=1}^{n} \hat{c_i} \implies \forall n, \sum_{i=1}^{n} a_i \ge 0$$

Key way of thinking:

Put the accounting cost on specific objects.

Accounting method: array doubling revisited

$$\hat{c}_i = 3 \textit{ vs. } \hat{c}_i = 2$$

$$\hat{c}_i = 3 = \underbrace{1}_{\text{insert}} + \underbrace{1}_{\text{move itself}} + \underbrace{1}_{\text{help move another}}$$

	$\hat{c}_i$	c_i (actual cost)	a_i (accounting cost)
INSERT (normal)	3	1	2
INSERT (expansion)	3	$1+t$	$-t+2$

Array merging (Problem 4.13): the summation method

$$\textsc{Create}\ (1);\ \textsc{Merge}\ (2m)$$

i	c_i
1	1
2	$1+2$
3	1
4	$1+2+4$
5	1
6	$1+2$
7	1
8	$1+2+4$
$\vdots$	$\ldots$

$$\sum_{i=1}^{n} c_i = \sum_{j=0}^{\lfloor \log n \rfloor} \lfloor \frac{n}{2^j} \rfloor 2^j \le n(\lfloor \log n \rfloor + 1)$$

$$\forall i, \hat{c}_i = 1 + \lfloor \log n \rfloor$$

Array merging (Problem 4.13): the accounting method

$$\hat{c}_i = 1 + \lfloor \log n \rfloor$$

What does it mean?

$$\forall n, \sum_{i=1}^{n} a_i \geq 0$$

Two stacks, one queue (Problem 4.14)

Algorithm 1 Simulating a queue using two stacks S_1, S_2.

procedure $\textsc{Enq}(x)$
    Push(S_1, x)

procedure $\textsc{Deq}()$
    if $S_2 = \emptyset$ **then**
        while $S_1 \neq \emptyset$ **do**
            Push$(S_2,$ *Pop*$(S_1))$
    Pop(S_2)

Two stacks, one queue: the summation method

$$(\sum_{i=1}^{n} c_i)/n$$

The operation sequence is NOT known.

Two stacks, one queue: the accounting method

item:	Push into S_1	Pop from S_1	Push into S_2	Pop from S_2
	1	1	1	1

$$\hat{c}_{\text{ENQ}} = 3$$
$$\hat{c}_{\text{DEQ}} = 1$$

$$\sum_{i=1}^{n} a_i \geq 0 \Longleftarrow \sum_{i=1}^{n} a_i = \#S_1 \times 2$$


THANK YOU

写清楚算法原理，不要只写代码

写清楚算法原理，不要只写代码

写清楚算法原理，不要只写代码