Algoritmi e Strutture Dati

Modelli di calcolo e analisi di algoritmi

Maria Rita Di Berardini[2], Emanuela Merelli[1]

[1]Scuola di Scienze e Tecnologie - Sezione di Informatica
Università di Camerino

[2]Polo di Scienze
Università di Camerino ad Ascoli Piceno

Analisi di Algoritmi

Analizzare un algoritmo vuol dire determinare le risorse necessarie all'algoritmo in termini di

spazio di memoria
(quantità di memoria utilizzata durante l'esecuzione)
e
tempo computazionale
(tempo di esecuzione)

Analisi di Algoritmi

L'analisi della complessità di un algoritmo in termini di tempo di esecuzione consente di:

- stimare il tempo impiegato
- stimare il più grande input gestibile in termini ragionevoli
- confrontare l'efficienza di due algoritmi diversi
- ottimizzare le parti "critiche"

Definiamo una funzione T : **dimensione** $\rightarrow$ **tempo** impiegato

Dimensione dell'input

Per molti problemi (ex. l'ordinamento) la misura più naturale è il numero di elementi (criteri di costo uniforme)

Per altri (ex. moltiplicazione di numeri interi) la misura migliore è il numero totale di bit necessari per la rappresentazione dell'input (criteri di costo logaritmico)

In realtà ciascun elemento è rappresentato da un numero costante di bit, quindi le due misure coincidono a meno di una costante moltiplicativa

In altri casi ancora, è più appropriato descrivere la dimensione con due numeri; ex: se l'input è una matrice bidimensionale la dimensione dell'input è $\#righe \times \#colonne$

Modello di Calcolo

Prima di analizzare un algoritmo abbiamo bisogno di stabilire quale sarà la tecnologia di riferimento utilizzata per eseguire gli algoritmi quando saranno realizzati come programmi

Assumiamo di utilizzare

Mono-Processore + RAM (Random Access Memory)

assenza di concorrenza e parallelismo

Definizione di tempo

Tempo = "wall-clock" time ossia il tempo tempo effettivamente impiegato per eseguire un algoritmo

Dipende da troppi fattori (non sempre prevedibili)

1. bravura del programmatore
2. linguaggio di programmazione utilizzato
3. processore, memoria (cache, primaria, secondaria)
4. sistema operativo, processi attualmente in esecuzione

Dobbiamo usare un modello astratto: introduciamo un concetto di tempo legato al "# di operazioni elementari" o di passi eseguiti per il calcolo dell'output corrispondente

Tempo di esecuzione

Numero di operazioni elementari o "passi" eseguiti per il calcolo dell'output

Numero di operazioni elementari o "passi" eseguiti per il calcolo dell'output

passo $\cong$ una linea di pseudocodice

Hp: ogni passo riferito ad una linea i, è eseguito in un tempo costante c_i

Parte I

Il problema dell'ordinamento

Il problema dell'ordinamento

Definition

Dato un insieme di n numeri $\langle a_1, a_2, \ldots, a_n \rangle$, trovare un'opportuna permutazione $\langle a'_1, a'_2, \ldots, a'_n \rangle$ tale che $a'_1 \leq a'_2 \leq \ldots \leq a'_n$

Input: $\langle a_1, a_2, \ldots, a_n \rangle$

Output: $\langle a'_1, a'_2, \ldots, a'_n \rangle$ oppure

$\langle a_{\pi(1)}, a_{\pi(2)}, \ldots, a_{\pi(n)} \rangle$

dove π è un'opportuna permutazione degli indici $1, \ldots, n$

Istanza del problema

Input: (31, 41, 59, 26, 41, 58)

Output: (26, 31, 41, 41, 58, 59)

La scelta del migliore algoritmo dipende:

- dal numero di elementi da ordinare
- da quanto gli elementi siano già ordinati
- dispositivo di memoria (metodo d'accesso)

Idea per ordinare

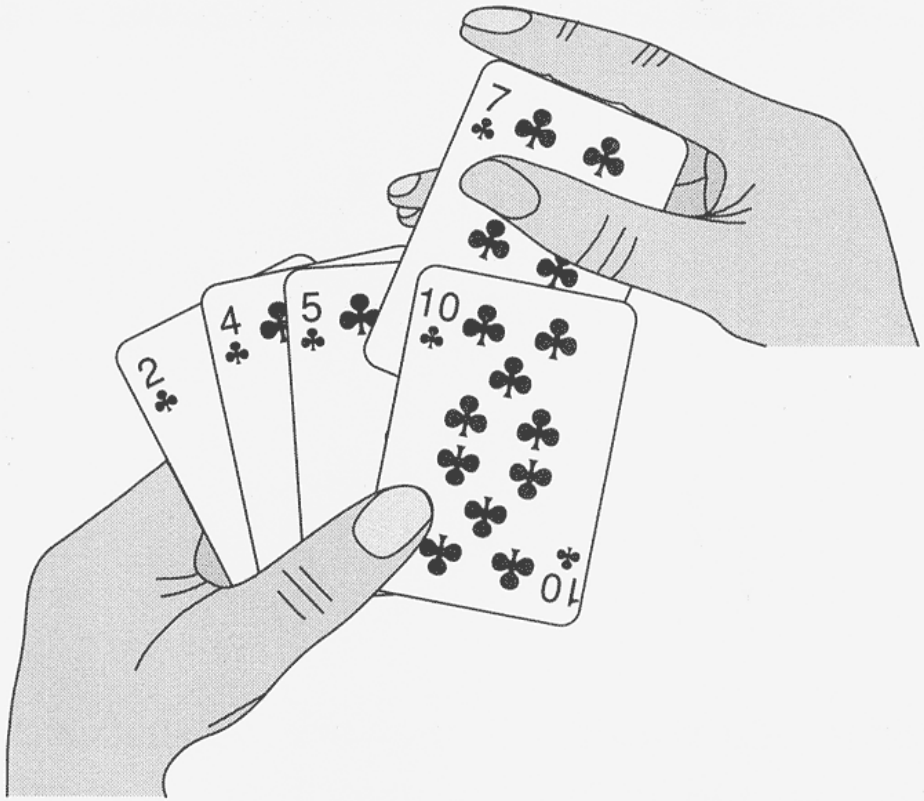
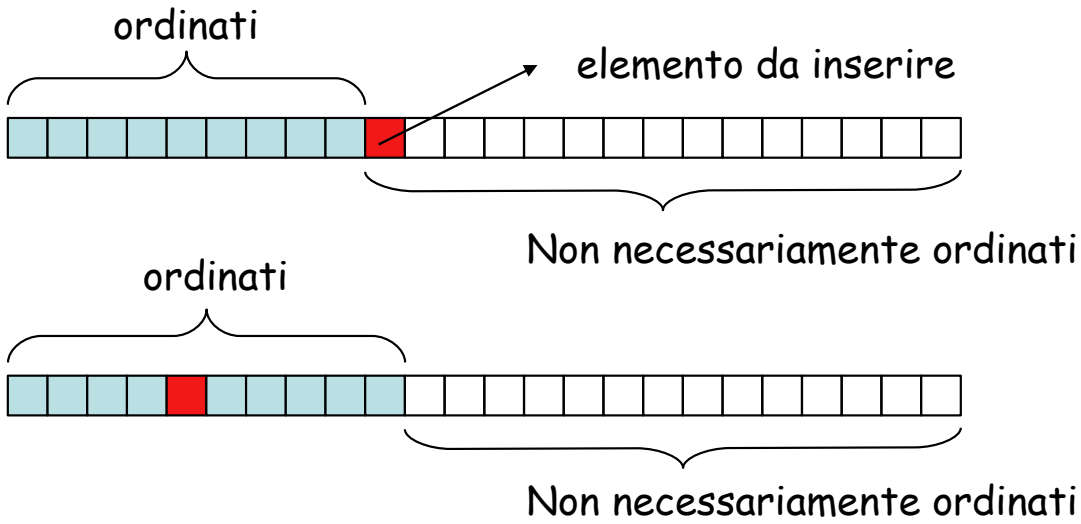


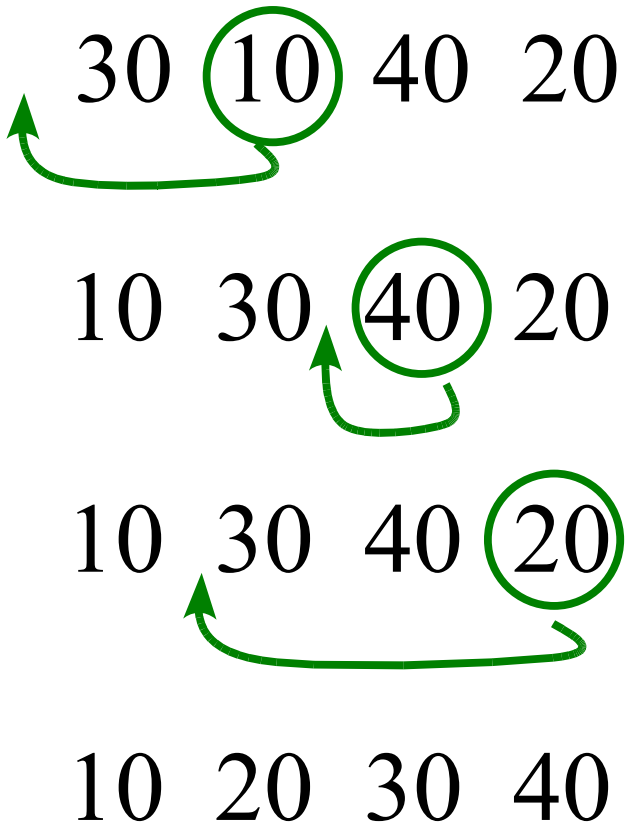
Figure 2.1 Sorting a hand of cards using insertion sort.

Idea per ordinare

Ad ogni passo si ha una sottosequenza ordinata in cui inserisco un nuovo elemento dell'input:



Idea per ordinare



Insertion Sort

L'algoritmo di ordinamento Insertion Sort risulta efficiente nel caso in cui il numero di elementi (n) da ordinare è piccolo

```
for j ← 2 to lenght[A]
    do key ← A[j]
        ▷ Si inserisce A[j] nella sequenza ordinata A[1..j-1]
        i ← j-1
        while i > 0 e A[i] > key
            do   A[i+1] ← A[i]
                 i ← i-1
        A[i+1] ← key
```

Istanza $A = \{5, 2, 4, 6, 1, 3\}$

$j = 2$	5	2	4	6	1	3
$j = 3$	2	5	4	6	1	3
$j = 4$	2	4	5	6	1	3
$j = 5$	2	4	5	6	1	3
$j = 6$	1	2	4	5	6	3
$j = 7$	1	2	3	4	5	6

Insertion Sort

Il tempo computazionale impiegato dalla procedura Insertion Sort dipende

- dalla dimensione input, il numero degli elementi
- dallo ordinamento implicito nella sequenza

Il tempo di esecuzione è espresso in funzione della dimensione del problema, cioè dell'input

Insertion Sort

Insertion-sort(A,n)

statement	**costo**
1. for j = 2 to n {	c_1
2. key = A[j]	c_2
3. i = j-1	c_3
4. while $(i > 0)$ and $(A[i] > key)$ {	c_4
5. $A[i+1] = A[i]$	c_5
6. $i = i - 1$	c_6
7. }	
8. $A[i+1] = key$	c_8
9. }	

Indichiamo con n_i il numero di volte che un passo i viene eseguito:

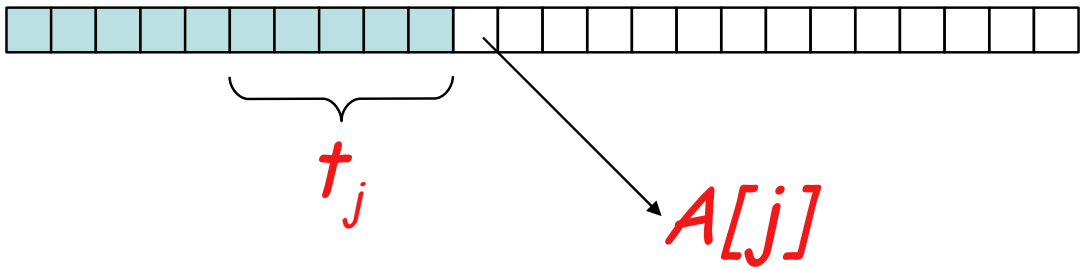
$$T(n) = n_1 c_1 + n_2 c_2 + \ldots n_k c_k = \sum_{i=1}^{k} n_i c_i$$

Insertion-sort

Insetion-sort(A,n)

statement	**costo**	**volte**
1. for j $=$ 2 to n $\{$	c_1	n
2. key $=$ A[j]	c_2	$n-1$
3. i $=$ j-1	c_3	$n-1$
4. while $(i > 0)$ and $(A[i] > key)$ $\{$	c_4	??
5. $A[i+1] = A[i]$	c_5	??
6. $i = i-1$	c_6	??
7. $\}$		
8. $A[i+1] = key$	c_8	$n-1$
9. $\}$		

Insertion Sort: analisi del costo computazionale



Fissato $j = 2, \ldots n$, denotiamo con t_j numero di elementi maggiori di $key = A[j]$ (t_j il numero di volte che il test del ciclo while viene eseguito, $t_j = t_j + 1$). Il valore di t_j dipende dal vettore in input:

- Caso ottimo (vettore già ordinato) $t_j = 0$ e $t_j = 1$
- Caso peggiore (ordinato in maniera decrescente) $t_j = j - 1$ e $t_j = j$
- Caso medio $t_j = j/2$

Insertion Sort: analisi del costo computazionale

Insetion-sort(A,n)

statement	**costo**	**volte**
1. for j = 2 to n {	c_1	n
2. key = A[j]	c_2	$n-1$
3. i = j-1	c_3	$n-1$
4. while $(i > 0)$ and $(A[i] > key)$ {	c_4	n_4
5. $A[i+1] = A[i]$	c_5	n_5
6. $i = i-1$	c_6	n_6
7. }		
8. $A[i+1] = key$	c_8	$n-1$
9. }		

$$n_4 = \sum_{j=2}^{n} t_j, \text{ e } n_5 = n_6 = \sum_{j=2}^{n}(t_j - 1)$$

Insertion Sort: analisi del costo computazionale

Insetion-sort(A,n)

statement	**costo**	**volte**
1. for j = 2 to n {	c_1	n
2. key = A[j]	c_2	$n-1$
3. i = j-1	c_3	$n-1$
4. while $(i > 0)$ and $(A[i] > key)$ {	c_4	$\sum_{j=2}^{n} t_j$
5. $A[i+1] = A[i]$	c_5	$\sum_{j=2}^{n}(t_j - 1)$
6. $i = i - 1$	c_6	$\sum_{j=2}^{n}(t_j - 1)$
7. }		
8. $A[i+1] = key$	c_8	$n-1$
9. }		

$$T(n) = c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 \sum_{j=2}^{n} t_j + (c_5 + c_6) \sum_{j=2}^{n} (t_j - 1)$$

Insertion Sort: analisi del costo computazionale nel caso ottimo

Nel **caso ottimo** $t_j = 1$, quindi

$$\begin{aligned} T(n) = & \quad c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 \sum_{j=2}^{n} t_j + (c_5 + c_6) \sum_{j=2}^{n} (t_j - 1) \\ = & \quad c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 \sum_{j=2}^{n} 1 \\ = & \quad c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 (n-1) \\ = & \quad (c_1 + c_2 + c_3 + c_4 + c_8) n - (c_2 + c_3 + c_4 + c_8) \end{aligned}$$

$$T(n) = an + b$$

dove a e b sono delle costanti

$T(n)$ è una funzione lineare in n

Insertion Sort: analisi del costo computazionale nel caso peggiore

Nel **caso peggiore** $t_j = j$, quindi

$$\begin{aligned} T(n) = & \quad c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 \sum_{j=2}^{n} t_j + (c_5 + c_6) \sum_{j=2}^{n} (t_j - 1) \\ = & \quad c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 \sum_{j=2}^{n} j + (c_5 + c_6) \sum_{j=2}^{n} (j - 1) \end{aligned}$$

Fact

$$\sum_{j=2}^{n} j = (\sum_{j=1}^{n} j) - 1 = \frac{n(n+1)}{2} - 1 = \frac{n^2}{2} + \frac{n}{2} - 1$$

$$\sum_{j=2}^{n} (j-1) = \sum_{j=1}^{n-1} j = \frac{(n-1)n}{2} = \frac{n^2}{2} - \frac{n}{2}$$

Insertion Sort: analisi del costo computazionale nel caso peggiore

$$
\begin{aligned}
T(n) = & \quad c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 \textstyle\sum_{j=2}^{n} j + (c_5 + c_6) \sum_{j=2}^{n} (j-1) \\
= & \quad c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 (\tfrac{n^2}{2} + \tfrac{n}{2} - 1) + (c_5 + c_6)(\tfrac{n^2}{2} - \tfrac{n}{2}) \\
= & \quad \tfrac{c_4 + c_5 + c_6}{2} n^2 + (c_1 + c_2 + c_3 + \tfrac{c_4 - c_5 - c_6}{2} + c_8) n - (c_2 + c_3 + c_4 + c_8)
\end{aligned}
$$

$$T(n) = an^2 + bn + c$$

dove a, b e c sono delle costanti

$T(n)$ è una funzione quadratica in n

Analisi nel caso medio

L'analisi del tempo di esecuzione nel **caso medio** analizza come si comporta l'algoritmo in "media"

È legato al fatto che in alcune circostanze non è necessario che un algoritmo operi con una quantità limitata di risorsa per ogni possibile input, ma è sufficiente che sia "generalmente efficiente"

Per ogni n, il costo dell'algoritmo viene calcolato come la somma dei costi per ogni possibile input di dimensione n tenendo conto della probabilità che ciascun singolo caso si verifichi

$$T_{medio}(n) = \sum_{\{I:|I|=n\}} T(I)Pr(I)$$

Analisi nel caso medio

Bisogna determinare la distribuzione di probabilità di ciascun input

Problema: non sempre si dispone delle informazioni necessarie per determinare la distribuzione Pr

Una possibile soluzione: fare delle assunzioni plausibili sulle caratteristiche del problema

Ad esempio, nel caso del Insersion sort è plausibile assumere che, mediamente, solo la metà degli elementi della porzione di vettore $A[1...j-1]$ sia maggiore di *key*, e quindi che $t_j = j/2$

Insertion Sort: analisi del costo computazionale nel caso medio

Nel **caso medio** $t_j = j/2$, quindi

$$\begin{aligned} T(n) = & \quad c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 \sum_{j=2}^{n} t_j + (c_5 + c_6) \sum_{j=2}^{n} (t_j - 1) \\ = & \quad c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 \sum_{j=2}^{n} j/2 \\ & + (c_5 + c_6) \sum_{j=2}^{n} (j/2 - 1) \end{aligned}$$

$$\sum_{j=2}^{n} j/2 = \frac{1}{2} \sum_{j=2}^{n} j = \frac{1}{2} (\sum_{j=1}^{n} j - 1) = \frac{1}{2} (\frac{n^2}{2} + \frac{n}{2} - 1) = \frac{n^2}{4} + \frac{n}{4} - \frac{1}{2}$$

$$\sum_{j=2}^{n} (j/2 - 1) = \sum_{j=2}^{n} \frac{j-2}{2} = \frac{1}{2} \sum_{j=2}^{n} j - 2 = \frac{1}{2} \sum_{j=0}^{n-2} j = \frac{1}{2} \sum_{j=1}^{n-2} j =$$

$$\frac{1}{2} \frac{(n-2)(n-1)}{2} = \frac{1}{2} \frac{n^2 - 3n + 2}{2} = \frac{n^2}{4} - \frac{3n}{4} + \frac{1}{2}$$

Insertion Sort: analisi del costo computazionale nel caso medio

$$
\begin{aligned}
T(n) = \quad & c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 \textstyle\sum_{j=2}^{n} j/2 \\
& + (c_5 + c_6) \textstyle\sum_{j=2}^{n} (j/2 - 1) \\
= \quad & c_1 n + (c_2 + c_3 + c_8)(n-1) + c_4 (\tfrac{n^2}{4} + \tfrac{n}{4} - \tfrac{1}{2}) + \\
& (c_5 + c_6)(\tfrac{n^2}{4} - \tfrac{3n}{4} + \tfrac{1}{2}) \\
= \quad & \tfrac{c_4 + c_5 + c_6}{4} n^2 + (c_1 + c_2 + c_3 + \tfrac{c_4 - 3c_5 - 3c_6}{2} + c_8) n + \\
& (\tfrac{c_5 + c_6}{2} - (c_2 + c_3 + \tfrac{c_4}{2} + c_8))
\end{aligned}
$$

$$T(n) = an^2 + bn + c$$

dove a, b e c sono delle costanti

$T(n)$ è una funzione quadratica in n

Analisi del caso peggiore vs. analisi del caso medio

Il tempo di esecuzione nel caso ottimo non è molto significativo

Di solito prenderemo in considerazione il tempo di esecuzione nel caso peggiore. Perchè??

- il tempo di esecuzione nel caso peggiore rappresenta una limitazione superiore sui tempi di esecuzione per qualsiasi input, non possiamo fare di peggio
- per alcuni algoritmi il caso peggiore si verifica abbastanza spesso (ex. ricerca di un elemento in una base di dati)
- il caso medio è in generale altrettanto cattivo quanto il caso peggiore (vedi Insertion Sort)