Lezione 6

ANALISI DELL'EFFICIENZA DEGLI ALGORITMI DI RICERCA NELLE TABELLE HASH A INDIRIZZAMENTO APERTO

In una tabella hash a indirizzamento aperto, ogni slot della tabella contiene un elemento del nostro universo delle chiavi (o del nostro sistema dinamico) oppure è vuoto.

Il fattore di carico non può essere più grande di 1

$$\boxed{\alpha = \frac{n}{m} \le 1}$$

perché $\underbrace{n}_{\text{\# di chiavi memorizzate}} \le \underbrace{m}_{\text{\# di slot della tabella}}$

Abbiamo un hashing di permutazione che definisce l'ordine di preferenza degli slot per ogni chiave.

$$h : \underset{\{k_1, \dots, k_n\}}{K} \longrightarrow S_m \quad (\text{"insieme delle permutazioni di } m \text{ elementi"})$$

$$h(k) = \langle \underbrace{h(k,0)}_{\text{posizione di prima scelta}}, \underbrace{h(k,1)}_{\text{di seconda scelta}}, \dots, \underbrace{h(k, m-1)}_{\text{di } m\text{-esima scelta}} \rangle \in S_m$$

Assumiamo che l'hashing di permutazione sia:

- UNIFORME: $\forall\ k,\ \forall\ \underset{\text{permutazione}}{\sigma} \in S_m \qquad Pr[h(k) = \sigma] = \dfrac{1}{m!}$

- INDIPENDENTE : $Pr[\,h(k_1)=\sigma_1 \mid h(k_2)=\sigma_2\,] = Pr[\,h(k_1)=\sigma_1\,]$

$$Pr[\,h(k_1)=\sigma_1 \wedge h(k_2)=\sigma_2\,] = Pr[\,h(k_1)=\sigma_1\,]\; Pr[\,h(k_2)=\sigma_2\,]$$

$\alpha \le 1 \iff$ ogni slot é occupato da al più una chiave.

Assumiamo nel seguito che nessuna cancellazione sia avvenuta.

6.1 ANALISI DELLA RICERCA SENZA SUCCESSO

THM: Sia data una tabella hash in cui le collisioni sono risolte mediante indirizzamento aperto, con fattore di carico $\alpha<1$.

Il numero atteso di prove degli slot é al più $\frac{1}{1-\alpha}$.

Quindi il tempo computazionale atteso é $O\left(\frac{1}{1-\alpha}\right)$.

PROF:

Ogni prova incontra uno slot occupato, eccetto l'ultima.

$T(n) = \Theta(1) + \mathbb{E}[X] = \mathbb{E}[X]$

X: numero di prove in una ricerca senza successo.

Allora
$$\underbrace{\mathbb{E}[X]}_{\text{numero atteso di prove}} = \sum_{i=1}^{\infty} Pr[X \ge i]$$

$\{X \ge i\}$ = abbiamo fatto almeno i prove.

$\forall i$, A_i: "L'i-esima prova é avvenuta e ha incontrato uno slot occupato."

$\{X \geq i\} = A_1 \cap A_2 \cap \dots \cap A_{i-1}$

$$Pr[X \geq i] = Pr[A_1 \cap \dots \cap A_{i-1}] = Pr[A_1] \cdot Pr[A_2 | A_1] \cdot Pr[A_3 | A_1 \cap A_2] \cdot \dots$$
$$\dots Pr[A_{i-1} | A_1 \cap \dots \cap A_{i-2}] =$$

Ora $\quad Pr[A_1] = \underbrace{\frac{n}{m}}_{\alpha}, \quad Pr[A_2 | A_1] = \frac{n-1}{m-1} < \overbrace{\frac{n}{m}}^{\alpha}, \dots$

$$Pr[A_{i-1} | A_1 \cap A_2 \cap \dots A_{i-2}] = \frac{n-(i-2)}{m-(i-2)} < \frac{n}{m} = \alpha$$

$$\leq \underset{\uparrow}{\overset{Pr[A_1]}{}}\alpha \cdot \alpha \cdots \underset{\uparrow}{\overset{Pr[A_{i-1}|A_1 \dots A_{i-2}]}{}}\alpha = \alpha^{i-1}. \qquad i \leq n+1$$

$$Pr[X \geq i] = 0 \qquad i > n+1$$

Quindi, $$E[X] = \sum_{i=1}^{\infty} Pr[X \geq i] = \sum_{i=1}^{n+1} Pr[X \geq i] + \overbrace{\sum_{i=n+2}^{\infty} Pr[X \geq i]}^{0} \leq$$
$$\sum_{i=1}^{n+1} \alpha^{i-1} = \sum_{i=1}^{\infty} \alpha^{i-1} = \sum_{i=0}^{\infty} \alpha^i \underset{\alpha<1}{=} \frac{1}{1-\alpha}. \quad \square$$

Oss: Se α è costante $\rightarrow E(\text{Time}) = O(1)$.

Esempio: Tabella hash mezza piena $\alpha = 1/2$, # prove in una ricerca senza successo $\leq \frac{1}{1-\frac{1}{2}} = 2$

Se la tabella è piena al 90%, $\alpha = \frac{90}{100} = \frac{9}{10} \rightarrow$ #prove $\leq \frac{1}{1-0.9} = 10$.

COR: Inserire un elemento in una tabella hash a indirizzamento aperto, $\alpha < 1$ richiede al più $\frac{1}{1-\alpha}$ prove in media.
Quindi $E(\text{Time}) = O\left(\frac{1}{1-\alpha}\right)$.

Proof: Ogni inserimento richiede una ricerca senza successo seguita dall'allocazione della chiave nel primo slot libero trovato. □

6.2 ANALISI DELLA RICERCA CON SUCCESSO

THM: Dato una tabella hash a indirizzamento aperto con fattore di carico $\alpha<1$, il numero atteso di prove in una ricerca con successo é al più $\frac{1}{\alpha}\ln\frac{1}{1-\alpha}$.

$$\mathbb{E}[\text{Time}] = O\left(\frac{1}{\alpha}\ln\frac{1}{1-\alpha}\right).$$

Assumiamo che l'hashing sia indipendente e uniforme e che nessuna cancellazione sia avvenuta.

PROOF: La ricerca di una chiave k riproduce la stessa sequenza di tentativi del suo inserimento.

Assumiamo che k sia la $(i+1)$-esima chiave inserita per qualche $i \in \{1, 2, \dots, n\}$.

Quando k è stato inserito il fattore di carico era $\tilde{\alpha} = \frac{i}{m}$ $\Rightarrow$ il numero atteso di tentativi per l'inserimento di k é $\frac{1}{1-\tilde{\alpha}} = \frac{1}{1-\frac{i}{m}} = \frac{m}{m-i}$.

$$\mathbb{E}[\#\text{ di prove}] = \sum_{i=0}^{n-1} \Pr[\text{stiamo cercando la chiave } k \text{ } (i+1)\text{-esima}] \cdot$$

$$\mathbb{E}[\#\text{ prove per cercare la chiave } (i+1)\text{-esima}]$$

$$= \sum_{i=0}^{n-1} \Pr[\text{stiamo cercando la chiave } (i+1)\text{-esima}] \cdot$$

$$\mathbb{E}[\#\ \text{prove per } \underline{\text{inserire}} \text{ la chiave } (i+1)\text{-esima}] =$$

$$= \sum_{i=0}^{n-1} \frac{1}{n} \cdot \frac{m}{m-i} = \frac{m}{n} \sum_{i=0}^{n-1} \frac{1}{m-i} = \frac{1}{\alpha} \underbrace{\sum_{j=m-n+1}^{m} \frac{1}{j}}_{(*)} \le$$

$j = m-i$

$i = m-j$

$i=0 \to j = m$

$i = n-1 \to j = m-n+1$

(*) Se $f(x)$ monotona e decrescente

$$\int_{a}^{b+1} f(x)\,dx \le \sum_{k=a}^{b} f(k) \le \int_{a-1}^{b} f(x)\,dx$$

$$\le \frac{1}{\alpha} \int_{m-n}^{m} \frac{1}{x}\,dx =$$

$$= \frac{1}{\alpha} \Big(\ln x \Big)_{m-n}^{m} =$$

$$= \frac{1}{\alpha} \left[\ln m - \ln (m-n) \right] =$$

$$= \frac{1}{\alpha} \ln \frac{m}{m-n} = \frac{1}{\alpha} \ln \frac{1}{1-\frac{n}{m}} =$$

$$= \frac{1}{\alpha} \ln \left(\frac{1}{1-\alpha} \right). \quad \square$$