Análisis amortizado: Expansión de tablas

Teoría de Algoritmos I (75.29 / 95.06)

Ing. Víctor Daniel Podberezski
vpodberezski@fi.uba.ar

Tablas (Lista, Hash...)

Diversas estructuras de datos utilizan vectores para almacenar datos.

La cantidad de datos a almacenar varia (aumentando o disminuyendo)

En ocasiones se requiere expandir (o contraer) el vector.

Este redimensión implica "mover" todos los datos a un nuevo vector.

Se debe definir en base al factor de carga

- Cuando expandirse
- Cuando contraerse

Primera aproximación (simplificada)

Se permitirá únicamente expandir la tabla

TABLE_INSERT(x)

Se expandirá:

Cuando el vector se llene

Al doble del tamaño actual

Definimos factor de carga a:

#T / size(T) donde size(T) capacidad de la tabla

y #T cantidad de elementos en la tabla.

TABLE_INSERT(x)

```
Si size(T)==0
    Crear T con size(T)=1

Si #T == size(T)
    Crear T’ con size(T’) = 2*size(T)
    Copiar T a T’
    Liberar T
    T = T’

insertar x en T
Incrementar #T
```

Inicialización

Expansión

Inserción

Costo de la inserción

El costo de la inserción i-esima es

- 1 si la tabla tiene lugar
- i si la tabla esta llena (copiar i-1 elementos + insertar elemento i)

La tabla se expande cuando i-1 es potencia de 2

Expansión de tablas - Aggregate analysis

Si insertamos n elementos tendremos:

$\lfloor \log_2 n \rfloor$ expansiones

n inserciones

El costo de las n operaciones:

$$\sum_{i=1}^{n} C_i \leq n + \sum_{i=1}^{\lfloor \log(n) \rfloor} 2^i \quad < n + 2n = 3n$$

El costo amortizado de cada operación:

$3 \rightarrow O(1)$

Expansión de tablas - Accounting method

Los costos de la operación "insertar" son variables

- Dependen de si se requiere expansión o no

Proponemos:

- Costo amortizado de 3 por cada inserción (1 de inserción + 2 "a cuenta").

Inicialmente el crédito es 0

Los costos de expansión se pagan con lo ahorrado por las inserciones

- 1 para moverse a si mismo
- 1 para mover otro elemento previo a la expansión anterior.

Expansión de tablas - Potential method

Definimos:

$\Phi(D_i) = 2*num(T) – size(T)$

Inicialmente $\Phi(D_0) = 0$

Al estar siempre la mitad o mas de la tabla ocupada → $\Phi(D_i) \geq 0$

Debemos analizar el costo amortizado para:

Inserción si no hay expansión

Inserción si hay expansión

Expansión de tablas - Potential method (cont.)

Si inserto el elemento i y no hay expansión:

$$\hat{C}_i = C_i + \Phi(D_i) - \Phi(D_{i-1}) \qquad C_i = 1$$

$$\Phi(D_i) = 2\,num(T_i) - \text{size}(T_i) \qquad \Phi(D_{i-1}) = 2\,num(T_{i-1}) - \text{size}(T_{i-1})$$

$$\text{size}(T_{i-1}) = \text{size}(T_i) \qquad num(T_i) = num(T_{i-1}) + 1$$

$$\hat{C}_i = 1 + [2\,num(T_i) - \cancel{\text{size}(T_i)}] - [2\,num(T_{i-1}) - \cancel{\text{size}(T_{i-1})}]$$

$$\hat{C}_i = 1 + 2[num(T_{i-1}) + 1] - 2\,num(T_{i-1})$$

$$\hat{C}_i = 3$$

Expansión de tablas - Potential method (cont.)

Si inserto el elemento i y hay expansión:

$$\hat{C}_i = C_i + \Phi(D_i) - \Phi(D_{i-1}) \qquad C_i = num(T_i)$$

$$\Phi(D_i) = 2\,num(T_i) - \text{size}(T_i) \qquad \Phi(D_{i-1}) = 2\,num(T_{i-1}) - \text{size}(T_{i-1})$$

$$2\,\text{size}(T_{i-1}) = \text{size}(T_i) \qquad num(T_i) = num(T_{i-1}) + 1$$

$$\text{size}(T_{i-1}) = num(T_{i-1})$$

$$\hat{C}_i = num(T_i) + [2\,num(T_i) - \text{size}(T_i)] - [2\,num(T_{i-1}) - \text{size}(T_{i-1})]$$

$$= num(T_i) + [2\,num(T_i) - 2\,\text{size}(T_{i-1})] - [2\,num(T_{i-1}) - (num(T_i - 1))]$$

$$= num(T_{i-1}) + 1 + [2(num(T_{i-1} + 1)) - 2\,\text{num}(T_{i-1})] - [2\,num(T_{i-1}) - (num(T_i - 1))]$$

$$\hat{C}_i = 3 \quad \Rightarrow \quad O(1)$$

Agregando la eliminación

Si permitimos la eliminación de elementos

- Se requiere contraer la tabla cuando el factor de carga sea menor a un valor
- Elegir incorrectamente el factor puede generar problemas (ej: ½)

Llamaremos al método

- TABLE_DELETE(x)

Proponemos

- Mantener el criterio de expansión
- Contraer la tabla a la mitad cuando el factor de carga sea menor a 1/4

Redimensión de tablas - Potential method

Definimos:

$$\Phi(D_i) = \begin{cases} 2\,num(T) - \text{size}(T) & \text{si factor carga } \geq 1/2 \\ \text{size}(T)/2 - num(T) & \text{si factor carga } < 1/2 \end{cases}$$

Inicialmente $\Phi(D_0) = 0$

La energía potencial siempre es mayor a cero.

Debemos analizar el costo amortizado para:

i-esa operación es Inserción

- Si factor de carga ≥ ½ y no hay expansión
- Si factor de carga ≥ ½ y hay expansión
- Si factor de carga es < ½ y luego de insertar el factor sigue por debajo de ½
- Si factor de carga es < ½ y luego de insertar el factor es ,mayor o igual a ½

Es valido el análisis del modelo simplificado

Redimensión de tablas - Potential method (cont.)

Debemos analizar el costo amortizado para (cont):

- I-esima operación es eliminación
  - Si factor de carga < ½ y luego de eliminar el factor es mayor a ¼ (sin contracción)
  - Si factor de carga < ½ y luego de eliminar el factor es menor o igual a ¼ (contracción)
  - Si factor de carga es ≥ ½ y luego de eliminar el factor sigue por arriba del ½
  - Si factor de carga es ≥ ½ y luego de eliminar el factor es menor a ½

Para cada uno de estos casos

- Determinar costo real, ecuación de potencial i-1 y ecuación de potencial i, equivalencias entre size y num de la tabla
- Calcular el costo amortizado (llegaremos a valores de 0,1,2,3)
- Por lo tanto podemos determinar que cada operación es general es O(1)

.UBA fiuba FACULTAD DE INGENIERÍA

Presentación realizada en Abril de 2020