Mảng cộng dồn (prefix sum) và ứng dụng

Trương Phước Hải

Khái niệm

- Mảng cộng dồn là CTDL lưu trữ tổng tích lũy của các phần tử trong một tập tính từ phần tử đầu tiên
- Mảng cộng dồn cho phép thực hiện hiệu quả thao tác tính tổng một nhóm các phần tử liên tiếp nhau

Mảng cộng dồn trên dãy

- Xét dãy các giá trị $a_1, a_2, \ldots, a_n$. Dãy các giá trị $s_0, s_1, \ldots, s_n$ được định nghĩa:
  - $s_0 = 0$
  - $s_i = a_1 + a_2 + \cdots + a_{i-1} + a_i$
- Khi đó dãy $s_0, s_1, \ldots, s_n$ được gọi là mảng cộng dồn (một chiều) của dãy $a_1, a_2, \ldots, a_n$

Mảng cộng dồn trên dãy

- Xây dựng mảng cộng dồn trên dãy

$$
\begin{aligned}
s_i &= a_1 + a_2 + \cdots + a_{i-1} + a_i \\
&= (a_1 + a_2 + \cdots + a_{i-1}) + a_i \\
&= s_{i-1} + a_i
\end{aligned}
$$

Mảng cộng dồn trên dãy

- Xây dựng mảng cộng dồn trên dãy

```
s[0] = 0;
for (i = 1; i <= n; ++i)
    s[i] = s[i-1] + a[i];
```

- Độ phức tạp của thao tác $O(n)$

Mảng cộng dồn trên bảng

- Xét bảng chữ nhật A gồm n dòng, m cột. Phần tử ở dòng i, cột j có giá trị $a[i][j] (1 \leq i \leq n; 1 \leq j \leq m)$.

A

5	-9	1	6	-8
2	3	-4	7	-10
-7	-12	4	1	-5
5	-6	-9	1	2

Mảng cộng dồn trên bảng

- Xét bảng chữ nhật S cùng kích thước với A, với $s[i][j]$ được xác định bởi công thức

$$s[i][j] = \sum_{u=1}^{i} \sum_{v=1}^{j} a[u][v]$$

A

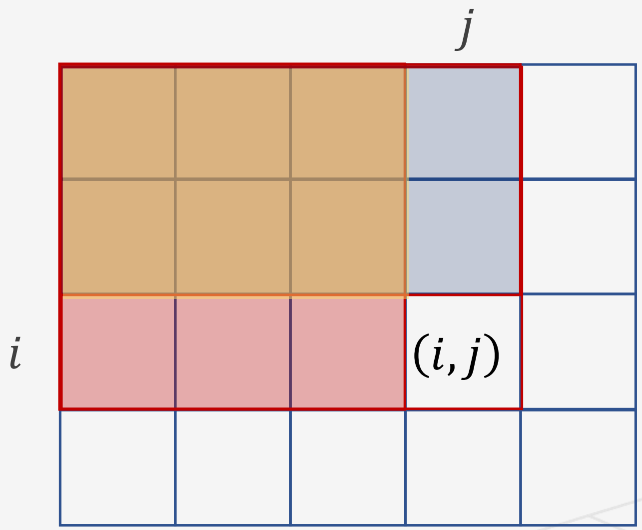
5	-9	1	6	-8
2	3	-4	7	-10
-7	-2	4	3	-3
5	6	2	-15	10

S

5	-4	-3	3	-5
7	1	-2	11	-7
0	-8	-7	9	-12
5	3	6	7	-4

Mảng cộng dồn trên bảng

- Bảng S được gọi là mảng cộng dồn 2 chiều của bảng A
- Xây dựng mảng cộng dồn 2 chiều



$$s[i][j] = s[i][j-1] + s[i-1][j] - s[i-1][j-1] + a[i][j]$$

Mảng cộng dồn trên bảng

- Xây dựng mảng cộng dồn trên bảng

```
s[i][j] = 0, ∀i,j
for (i = 1; i <= n; ++i)
    for (j = 1; j <= m; ++j)
        s[i][j] = s[i][j-1] + s[i-1][j]
                - s[i-1][j-1] + a[i][j];
```

- Độ phức tạp của thao tác $O(n \times m)$

Áp dụng 1

- Cho dãy gồm n giá trị $a_1, a_2, \dots, a_n$ và m truy vấn có dạng $[l, r]$ yêu cầu trả về giá trị $a_l + a_{l+1} + \cdots + a_r$

1	2	3	4	5	6	7	8	9	10
5	-10	2	7	-6	-4	8	-3	5	7
		l					r		

Thuật toán tầm thường

- Với mỗi truy vấn $[l, r]$, duyệt qua đoạn phần tử và tính tổng của chúng

```
for (i = 0; i < m; ++i) {
    Sum = 0;
    for (j = l[i]; j <= r[i]; ++j)
        Sum = Sum + a[j];
    output Sum;
}
```

Thuật toán tầm thường

- Đánh giá thuật toán
  - Độ phức tạp của mỗi truy vấn $O(n)$
  - Độ phức tạp trả lời m truy vấn $O(m \times n)$
- Thao tác xét và in kết quả của từng truy vấn là không thể cải tiến. Tìm cách cải tiến thao tác tính kết quả của từng truy vấn

Phương pháp mảng cộng dồn

- Nhận xét tổng các phần tử trong đoạn $[l, r]$

$$Sum = a_l + a_{l+1} + \cdots + a_r$$

$$= a_1 + a_2 + \cdots + a_{l-1} + a_l + \cdots + a_r - (a_1 + a_2 + \cdots + a_{l-1})$$

$$= s_r - s_{l-1}$$

Phương pháp mảng cộng dồn

- Cải tiến thao tác trả lời truy vấn với mảng cộng dồn

```
for (i = 0; i < m; ++i) {
    Sum = s[r[i]] – s[l[i]-1];
    output Sum;
}
```

- Sử dụng mảng cộng dồn giúp độ phức tạp của thao tác trả lời một truy vấn giảm xuống còn $O(1)$

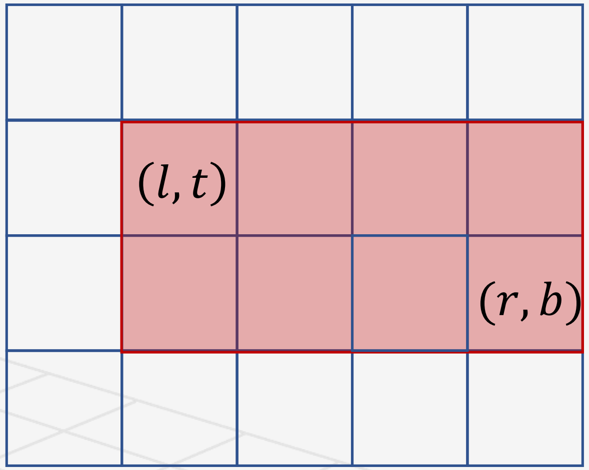
Phương pháp mảng cộng dồn

- Phương pháp thực hiện qua 2 công đoạn
  - Xây dựng mảng cộng dồn, độ phức tạp $O(n)$
  - Trả lời m truy vấn, độ phức tạp $O(m)$
- Độ phức tạp của thuật toán $O(n+m)$

Áp dụng 2

- Cho bảng chữ nhật kích thước n dòng, m cột. Phần tử ở dòng i, cột j có giá trị $a[i][j]$. Yêu cầu trả lời q truy vấn có dạng $[l, t, r, b]$ cho biết giá trị của biểu thức

$$Sum = \sum_{x=l}^{r} \sum_{y=t}^{b} a[x][y]$$



Ý tưởng chung

- Trả lời cho m truy vấn

```
for (i = 0; i < q; ++i) {
    Sum = SumRect(l[i], t[i], r[i], b[i]);
    output Sum;
}
```

- Độ phức tạp $O(q \times T)$, với T là thời gian để thực hiện một truy vấn

Thuật toán tầm thường

- Duyệt qua tất cả phần tử trong vùng chữ nhật xác định bởi 2 góc (l, t) và (r, b) để tính tổng các phần tử

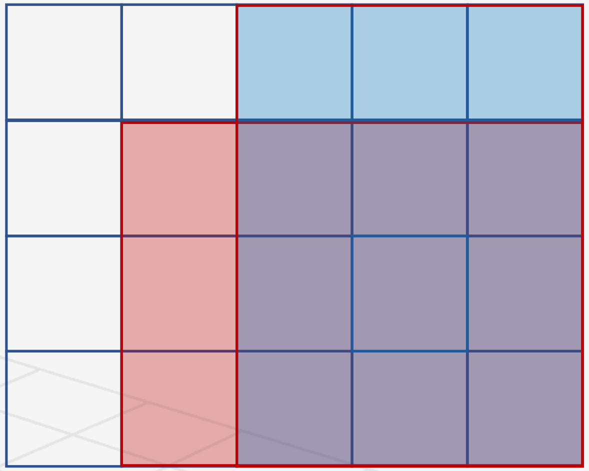
```
SumRect(l, t, r, b) {
    Sum = 0;
    for (i = l; i <= r; ++i)
        for (j = t; j <= b; ++j)
            Sum = Sum + a[i][j];
    return Sum;
}
```

Thuật toán tầm thường

- Đánh giá phương pháp
  - Độ phức tạp của mỗi truy vấn: $O(n \times m)$
  - Độ phức tạp của bài toán: $O(q \times n \times m)$

Phương pháp mảng cộng dồn

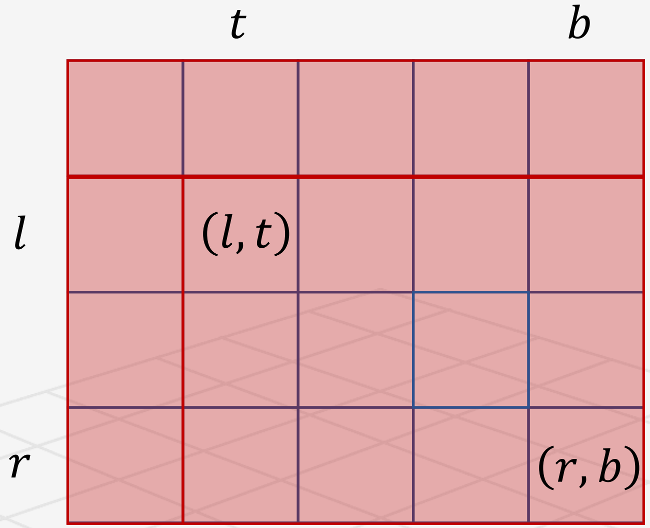
- Nhận xét
  - Phương pháp tầm thường thiếu hiệu quả do các phần tử thuộc một số vùng chữ nhật (thuộc phần giao) bị duyệt nhiều lần



Phương pháp mảng cộng dồn

- Sử dụng mảng cộng dồn để tính tổng các phần tử trong vùng hình chữ nhật

$$Sum = s[r][b] - s[l-1][b] - s[r][t-1] + s[l-1][t-1]$$



Phương pháp mảng cộng dồn

- Phương pháp thực hiện qua 2 công đoạn
  - Xây dựng mảng cộng dồn 2 chiều, độ phức tạp $O(n \times m)$
  - Trả lời q truy vấn, độ phức tạp $O(q)$
- Độ phức tạp của thuật toán: $O(n \times m)$

Bài tập áp dụng

- Bài 1: Cho dãy số $a_0, a_1, \dots, a_{n-1}$. Tìm một đoạn con có tổng các phần tử là lớn nhất
- Bài 2: Cho dãy số $a_0, a_1, \dots, a_{n-1}$. Tìm một đoạn con dài nhất có tổng các phần tử bằng 0
- Bài 3: Cho dãy số $a_0, a_1, \dots, a_{n-1}$. Đếm số đoạn con có tổng các phần tử bằng 0

Bài tập áp dụng

- Bài 4: Cho dãy số $a_0, a_1, \dots, a_{n-1}$. Tìm một đoạn con gồm ít nhất k phần tử sao cho tổng của chúng là lớn nhất
- Bài 5: Đặt dãy số không âm $a_0, a_1, \dots, a_{n-1}$ trên vòng tròn theo chiều kim đồng hồ. Tìm một đoạn con ngắn nhất có tổng là x
- Bài 6: Cho dãy số nguyên $a_0, a_1, \dots, a_{n-1}$. Tìm một đoạn con có tổng các phần tử chia dư cho m đạt lớn nhất

Bài tập áp dụng

- Bài 7: Cho bảng gồm n dòng, m cột. Phần tử ở dòng i, cột j mang giá trị a_{ij}. Tìm một vùng hình vuông con lớn nhất của bảng chỉ gồm các số chính phương
- Bài 8: Cho bảng gồm n dòng, m cột. Phần tử ở dòng i, cột j mang giá trị a_{ij}. Tìm một vùng chữ nhật con của bảng có tổng các phần tử là lớn nhất.

Bài tập áp dụng

- Bài 9: Cho dãy số $a_0, a_1, \dots, a_{n-1}$. Tìm cách chia dãy thành 2 đoạn sao cho tích của tổng 2 đoạn là lớn nhất