

Bài Toán 6: Tree Edit Distance

Đề Án 4: Duyệt Đồ Thị & Cây

Phát biểu bài toán

Cho hai cây có gốc T_1 và T_2 . Tính số phép biến đổi tối thiểu cần thiết để biến T_1 thành T_2 bằng ba thao tác:

- Insert (thêm nút)
- Delete (xóa nút)
- Rename (đổi nhãn)

Mục tiêu

Tìm khoảng cách edit $TED(T_1, T_2)$ sao cho tổng chi phí nhỏ nhất.

Thuật toán áp dụng

- (a) **Brute-force / Backtracking:** thử tất cả phép biến đổi
- (b) **Branch-and-bound:** loại bỏ nhánh có chi phí tạm tính $>$ best hiện tại
- (c) **Divide and Conquer:** chia cây thành các subtree nhỏ
- (d) **Dynamic Programming:** giải bài toán con bằng quy hoạch động

Chi tiết thuật toán (Zhang Shasha)

- Định nghĩa post-order cho mỗi node i trong cây
- Hàm $\text{ted}(i, j)$ = chi phí biến subtree rooted tại i thành subtree rooted tại j
- Dựng bảng $\text{dp}[i][j]$ với công thức:

$$\text{dp}[i][j] = \begin{cases} \text{dp}[i-1][j-1] & \text{nếu } \text{label}(i) = \text{label}(j) \\ \min \begin{cases} \text{dp}[i-1][j] + \text{cost}(\text{delete}) \\ \text{dp}[i][j-1] + \text{cost}(\text{insert}) \\ \text{dp}[i-1][j-1] + \text{cost}(\text{rename}) \end{cases} & \text{trường hợp khác} \end{cases}$$

Chú thích các biến số

- T_1, T_2 : hai cây gốc
- $dp[i][j]$: khoảng cách edit từ subtree i của T_1 tới subtree j của T_2
- $label[i]$: nhãn (tên) của node i
- $l[i], r[i]$: chỉ số preorder hoặc postorder của node i
- $cost_ins, cost_del, cost_ren$: chi phí các thao tác