

Project I

Các thuật toán nén nguồn rời rạc

Lê Minh Việt Anh 20210037¹, Nguyễn Trọng Nhật 20215625², Lê Việt Quang 20215630³, and Bùi Mạnh Chiến 20210118⁴

^{1,2,3}Kỹ thuật máy tính 05 - Khóa 66

⁴Kỹ thuật máy tính 03 - Khóa 66

Ngày 24 tháng 12 năm 2023

Đề bài: Thuật toán mã hóa của Huffman

Cho nguồn tin $X = \{x_1, x_2, \dots, x_q\}$ có phân bố xác suất $P(X) = \{p_1, p_2, \dots, p_q\}$ được nhập từ bàn phím. Thuật toán tìm mã Huffman với cơ số mã r có các ký hiệu mã có giá trị từ 0 đến $r - 1$ gồm các bước (tìm các từ mã ứng với mỗi tin của nguồn):

1. Tính giá trị $a = \frac{q-r}{r-1}$. Nếu a không nguyên thì tính giá trị mới $a' = \text{int}(a) + 1$. Tiếp theo tính số tin mới của nguồn $q' = r + a'(r-1)$. Số tin ảo có xác suất bằng 0 là $q' - q$. Nếu a nguyên thì $q' = q$.
2. Xếp các tin của nguồn q' theo thứ tự xác suất giảm dần.
3. Nhóm r tin cuối của nguồn thành một tin phụ có xác suất bằng tổng các xác suất của các tin được nhóm vào. Đánh dấu mỗi tin được nhóm bằng một ký hiệu mã khác nhau.
4. Lặp lại các bước 2 và 3 cho đến khi nguồn còn đúng r tin. Đánh dấu tin của nguồn này bằng một ký hiệu mã khác nhau.
5. Từ mã ứng với mỗi tin là chuỗi ký hiệu mã dùng để đánh dấu chính tin này và các tin phụ chứa nó theo thứ tự ngược với thứ tự khi nhóm.

Xây dựng chương trình cho phép:

1. Tìm bộ mã (từ mã) cho mỗi tin của nguồn.
2. Tạo bản tin (chuỗi tin) từ nguồn bằng bàn phím và sử dụng bộ mã vừa tìm được để mã hóa bản tin, tạo thành bản mã (chuỗi các ký hiệu mã) và sau đó từ bản tin giải mã lại bản tin.

1 Phân tích bài toán

Đề bài yêu cầu tạo mã Huffman cho nguồn nhập từ bàn phím với xác suất và cơ số bất kỳ, sau đó mã hóa một đoạn văn bản và giải mã chính đoạn mã đó. Ba công việc này có thể hoàn thành một cách độc lập.

1.1 Tạo mã Huffman cho nguồn nhập từ bàn phím

| | |
|---------|--|
| Đầu vào | Các cặp giá trị (nguồn, xác suất) với tổng xác suất bằng 1 |
| Đầu ra | Mã Huffman cho mỗi tin của nguồn theo cơ số cho trước |

1.2 Mã hóa bản tin

1.3 Giải bản mã