

# Bài 2 Ví dụ về tốc độ ngôn ngữ và độ dư thừa

Nhân Nguyễn Văn

April 2025

## Giải thích khái niệm và ví dụ minh họa

### 1. Công thức tốc độ ngôn ngữ:

$$R(X) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n)$$

Công thức này tính tốc độ ngôn ngữ khi chuỗi có số lượng ký tự rất lớn (hướng tới vô hạn).

### 2. Tốc độ ngôn ngữ tuyệt đối (khi các ký tự xuất hiện đều):

$$R(X) = \log_2(L)$$

Với  $L$  là số lượng ký tự khác nhau. Trong trường hợp này, không cần dùng đến giới hạn  $\lim_{n \rightarrow \infty}$  vì chỉ xét một chuỗi có độ dài hữu hạn.

### 3. Độ dư thừa thông tin:

$$D = R(X) - r(X)$$

Trong đó:

- $R(X)$ : tốc độ tối đa (entropy lý thuyết)
- $r(X)$ : tốc độ thực tế (entropy quan sát được)
- $D$ : phần thông tin bị dư do trùng lặp, lặp lại

## Ví dụ 1: Ngôn ngữ phân bố đều

- Bảng chữ cái:  $\{A, B, C, D\}$
- Mỗi ký tự có xác suất xuất hiện  $p = 0.25$
- Entropy thực tế:

$$r(X) = - \sum_{i=1}^4 0.25 \cdot \log_2(0.25) = 4 \cdot 0.25 \cdot 2 = 2 \text{ bits}$$

- Tốc độ tối đa:  $R(X) = \log_2(4) = 2$
- Dư thừa:  $D = 2 - 2 = 0$  (không dư thừa)

## Ví dụ 2: Ngôn ngữ phân bố lệch

- Phân bố: A (80%), B (10%), C (5%), D (5%)
- Tính gần đúng:

$$r(X) \approx -[0.8 \log_2(0.8) + 0.1 \log_2(0.1) + 2 \cdot 0.05 \log_2(0.05)] \approx 1.02 \text{ bits}$$

- Tốc độ tối đa:  $R(X) = \log_2(4) = 2$
- Dư thừa:  $D = 2 - 1.02 = 0.98 \text{ bits}$

## Kết luận

- Entropy càng thấp  $\rightarrow$  ngôn ngữ càng dư thừa  $\rightarrow$  tốc độ truyền tin càng kém hiệu quả.
- Ngôn ngữ đều đặn, không lặp  $\rightarrow$  tốc độ cao, không dư thừa.
- Ngôn ngữ lệch, nhiều lặp  $\rightarrow$  tốc độ thấp, dư thừa cao.