

Phương pháp đánh giá vòng sơ tuyển (Quy chế đánh giá chính thức cho mỗi câu truy vấn)

BTC Hội thi Thử thách trí tuệ nhân tạo AI Challenge TP. Hồ Chí Minh năm 2025

Ngày 22 tháng 8 năm 2025

1 Tổng quan

Đối với mỗi truy vấn, thí sinh được gửi tới đa 100 câu trả lời. Điểm cuối cùng được tính dựa trên chỉ số **Mean of Top-k R-Scores**, là trung bình cộng của các điểm tương quan cao nhất tại các ngưỡng xếp hạng (k) khác nhau.

Công thức tính điểm cuối cùng cho một truy vấn:

$$\text{Final Score} = \frac{1}{5} \sum_{k \in \{1, 5, 20, 50, 100\}} \left(\max_{1 \leq i \leq k} \{ \text{R-Score}(r_i) \} \right)$$

Trong đó:

- r_i : là câu trả lời ở vị trí xếp hạng thứ i .
- k : là các ngưỡng xếp hạng, bao gồm $\{1, 5, 20, 50, 100\}$.
- $\text{R-Score}(r_i)$: là điểm tương quan của một câu trả lời đơn lẻ, được tính khác nhau tùy theo từng nhiệm vụ.

2 Cách tính Điểm Tương Quan (R-Score)

Điểm tương quan $\text{R-Score}(r_i)$ đo lường mức độ chính xác của một câu trả lời r_i so với đáp án (Ground Truth - GT).

2.1 Dạng truy vấn 1: Textual-KIS

- **Định dạng trả lời (r_i):** <Tên file video>, <Frame Idx>
- **Điều kiện:** Một câu trả lời được xem là chính xác nếu khớp **tên file video** ($v_i = GT_v$) và **frame index** nằm trong khoảng đáp án đúng ($id_i \in [s, e]$).
- **Công thức:**

$$\text{R-Score}(r_i) = \mathbb{I}(v_i = GT_v \wedge id_i \in [s, e])$$

Trong đó $\mathbb{I}(\cdot)$ là hàm chỉ thị, trả về 1 nếu điều kiện đúng và 0 nếu sai.

2.2 Dạng truy vấn 2: Visual Question Answering

- **Định dạng trả lời (r_i):** <Tên file video>, <Frame Idx>, <Answer>
- **Điều kiện:** Một câu trả lời được xem là chính xác nếu khớp **tên file video** ($v_i = GT_v$), **frame index** nằm trong khoảng đáp án đúng ($id_i \in [s, e]$), và **câu trả lời** khớp với đáp án ($a_i = GT_a$).
- **Công thức:**

$$\text{R-Score}(r_i) = \mathbb{I}(v_i = GT_v \wedge id_i \in [s, e] \wedge a_i = GT_a)$$

2.3 Dạng truy vấn 3: Temporal-alignment

- **Định dạng trả lời (r_i):** <Tên file video>, <Frame ID_1>, ..., <Frame ID_N>
- **Điều kiện:** Câu trả lời phải khớp **tên file video** ($v_i = GT_v$). Nếu không, điểm sẽ là 0.
- **Công thức:** Nếu khớp tên video, điểm được tính bằng tỉ lệ các frame được gửi khớp với frame GT trong bán kính nhất định $[s, e]$.

$$\text{R-Score}(r_i) = \begin{cases} \frac{1}{N} \sum_{j=1}^N \mathbb{I}(id_{i,j} \in [s_j, e_j]) & \text{nếu } v_i = GT_v \\ 0 & \text{nếu } v_i \neq GT_v \end{cases}$$

Trong đó N là tổng số khoảng khắc trong truy vấn.

3 Cách tính Điểm Cuối Cùng

1. **Tính Top-k R-Score ($R@k$):** Với mỗi ngưỡng k , tìm điểm tương quan cao nhất trong top k câu trả lời đầu tiên.

$$R@k = \max_{1 \leq i \leq k} \{\text{R-Score}(r_i)\}$$

2. **Tính Final Score:** Lấy trung bình cộng của các điểm $R@k$ tại 5 ngưỡng.

$$\text{Final Score} = \frac{1}{5} \sum_{k \in \{1, 5, 20, 50, 100\}} R@k$$