

Tên: Nguyễn Thị Huyền Trang

MSSV: 21520488

Bài tập 4: BÀI TẬP KHUYẾN NGHỊ DỰA TRÊN NỘI DUNG

- Tiền xử lý dữ liệu: Loại bỏ từ trùng
dict = {cuộc, chiến_đầu, chống, lại, linh_hồn, ác_quỷ, đột_nhập, của, Thom, vào, một, ngân_hàng, tại, Hồng_Kông, những, rắc_rối, hài_hước, đòi_thường, cô_gái, 30, siêu, anh_hùng, cùng, mỗi, nguy_hiểm}

Câu 1: Chuyển tất cả các bộ phim về vector Boolean và sử dụng cosine để tìm ra top 1 bộ phim khuyến nghị cho user1.

- Vector Boolean:

M1 = {1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0}

M2 = {1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0}

M3 = {0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0}

M4 = {0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1}

M5 = {0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0}

- Sử dụng cosine để tìm ra top 1 bộ phim khuyến nghị cho user1

$$\cos(\vec{M}_5, \vec{M}_1) = \frac{0}{\sqrt{3} \cdot \sqrt{6}} = 0$$

$$\cos(\vec{M}_5, \vec{M}_2) = \frac{3}{\sqrt{3} \cdot \sqrt{9}} = \frac{\sqrt{3}}{9}$$

$$\cos(\vec{M}_5, \vec{M}_3) = \frac{3}{\sqrt{3} \cdot \sqrt{7}} = \frac{\sqrt{21}}{7}$$

$$\cos(\vec{M}_5, \vec{M}_4) = \frac{0}{\sqrt{3} \cdot \sqrt{7}} = 0$$

Vậy top 1 bộ phim khuyến nghị cho user1 là bộ phim có ID = M3

Câu 2: Chuyển tất cả các bộ phim về vector TF-IDF và sử dụng cosine để tìm ra top 1 bộ phim khuyến nghị user1.

- Vector TF-IDF:

M1 = {0.018, 0.043, 0.018, 0.018, 0.043, 0.043, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0}

M2 = {0.014, 0, 0, 0, 0, 0, 0.033, 0.014, 0.033, 0.033, 0.033, 0.033, 0.033, 0.033, 0, 0, 0, 0, 0, 0, 0, 0}

M3 = {0, 0, 0, 0, 0, 0, 0, 0.018, 0, 0, 0, 0, 0, 0, 0.043, 0.043, 0.043, 0.043, 0.043, 0, 0, 0, 0, 0}

M4 = {0, 0, 0.018, 0.018, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.043, 0.043, 0.043, 0.043, 0.043}

M5 = {0, 0, 0, 0, 0, 0, 0, 0.032, 0, 0, 0, 0, 0, 0, 0, 0, 0.074, 0, 0.074, 0, 0, 0, 0, 0}

- Sử dụng cosine để tìm ra top 1 bộ phim khuyến nghị cho user1

$$\cos(\vec{M}_5, \vec{M}_1) = 0$$

$$\cos(\vec{M}_5, \vec{M}_2) = 0,046$$

$$\cos(\vec{M}_5, \vec{M}_3) = 0,593$$

$$\cos(\vec{M}_5, \vec{M}_4) = 0$$

Vậy top 1 bộ phim khuyến nghị cho user1 là bộ phim có ID = M3

Câu 3: Sử dụng độ đo Jaccard để tìm ra top 2 bộ phim khuyến nghị user1.

$$J(M_5, M_1) = \frac{0}{11} = 0$$

$$J(M_5, M_2) = \frac{1}{14-1} = \frac{1}{13}$$

$$J(M_5, M_3) = \frac{3}{12-3} = \frac{1}{3}$$

$$J(M_5, M_4) = \frac{0}{12} = 0$$

Vậy top 2 bộ phim khuyến nghị cho user1 là bộ phim có ID là M2 và M3