# DS300.N11 - Hệ Khuyến Nghị
# Bài tập khuyến nghị dựa trên nội dung

Nguyễn Văn Kiệt, Huỳnh Văn Tín
Sinh viên: Phạm Đức Thể - 19522253

Thứ 3, ngày 27 tháng 09 năm 2022

## Bài tập

Khuyến nghị dựa trên nội dung

**Bộ dữ liệu phim**

| ID phim | Nội dung chính |
|---|---|
| M1 | Cuộc chiến đấu chống lại linh hồn ác quỷ |
| M2 | Cuộc đột nhập của Thom vào một ngân hàng tại Hồng Kông |
| M3 | Những rắc rối hài hước đời thường của cô gái 30 |
| M4 | Siêu anh hùng cùng chống lại mối nguy hiểm |

**Lịch sử xem phim của User1**

| Lịch sử xem phim | ID phim | Nội dung chính |
|---|---|---|
| User1 | M5 | Tình yêu hài hước của cô gái biên kịch |

**Tiền xử lý dữ liệu**

- Tách từ

| ID phim | Nội dung chính |
|---|---|
| M1 | Cuộc chiến_đấu chống lại linh_hồn ác_quỷ |
| M2 | Cuộc đột_nhập của Thom vào một ngân_hàng tại Hồng_Kông |
| M3 | Những rắc_rối hài_hước đời_thường của cô_gái 30 |
| M4 | Siêu anh_hùng cùng chống lại mối nguy_hiểm |

| Lịch sử xem phim | ID phim | Nội dung chính |
|---|---|---|
| User1 | M5 | Tình_yêu hài_hước của cô_gái biên_kịch |

- Chuyển nội dung về từ thường

| ID phim | Nội dung chính |
|---|---|
| M1 | cuộc chiến_đấu chống lại linh_hồn ác_quỷ |
| M2 | cuộc đột_nhập của Thom vào một ngân_hàng tại Hồng_Kông |
| M3 | những rắc_rối hài_hước đời_thường của cô_gái 30 |
| M4 | siêu anh_hùng cùng chống lại mối nguy_hiểm |

| Lịch sử xem phim | ID phim | Nội dung chính |
|---|---|---|
| User1 | M5 | tình_yêu hài_hước của cô_gái biên_kịch |

- Loại bỏ từ trùng
  Từ điển = {cuộc, chiến_đấu, chống, lại, linh_hồn, ác_quỷ, đột_nhập, của, Thom, vào, một, ngân_hàng, tại, Hồng_Kông, những, rắc_rối, hài_hước, đời_thường, cô_gái, 30, siêu, anh_hùng, cùng, mối, nguy_hiểm}

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, 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, 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, 0, 0}
     - M4 = {0, 0, 1, 1, 0, 0, 0, 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, 0, 0}
   - Tính toán sử dụng cosine:
     - Công thức:

$$cos(\vec{a}, \vec{b}) = \frac{\vec{a}.\vec{b}}{|\vec{a}|.|\vec{b}|} = \frac{a_1.b_1 + a_2.b_2}{\sqrt{a_1^2 + a_2^2}.\sqrt{b_1^2 + b_2^2}}$$

     - Áp dụng:

$$cos(\vec{M}5, \vec{M}1) = \frac{\vec{M}5.\vec{M}1}{|\vec{M}5|.|\vec{M}1|} = \frac{0}{\sqrt{3}.\sqrt{6}} = 0$$

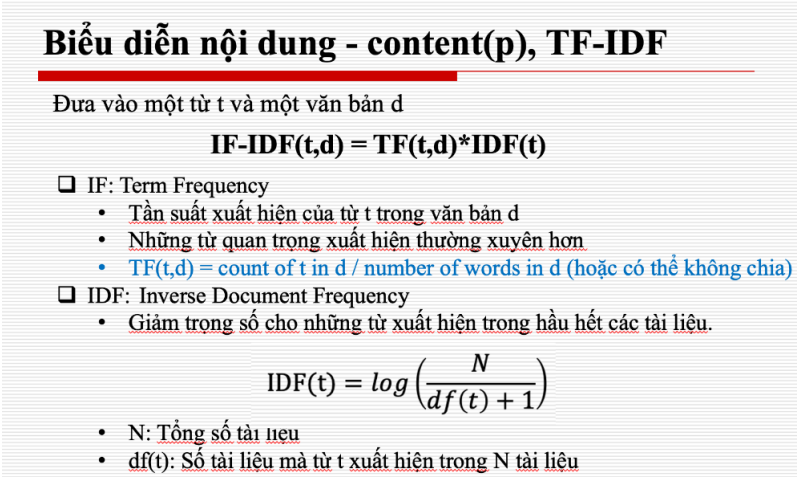
$$cos(\vec{M}5, \vec{M}2) = \frac{\vec{M}5.\vec{M}2}{|\vec{M}5|.|\vec{M}2|} = \frac{1}{\sqrt{3}.\sqrt{9}} = \frac{\sqrt{3}}{9}$$

$$cos(\vec{M}5, \vec{M}3) = \frac{\vec{M}5.\vec{M}3}{|\vec{M}5|.|\vec{M}3|} = \frac{3}{\sqrt{3}.\sqrt{7}} = \frac{\sqrt{21}}{7}$$

$$cos(\vec{M}5, \vec{M}4) = \frac{\vec{M}5.\vec{M}4}{|\vec{M}5|.|\vec{M}4|} = \frac{0}{\sqrt{3}.\sqrt{7}} = 0$$

   - Kết luận: Top 1 bộ phim khuyến nghị cho user1 là bộ phim có ID là M3.

2. Chuyển tất cả các bộ phim về vector TF-IDF và sử dùng 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, 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, 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, 0.043, 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, 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, 0}
   - Tính toán sử dụng cosine:

– Công thức:

$$cos(\vec{a},\vec{b}) = \frac{\vec{a}.\vec{b}}{|\vec{a}|.|\vec{b}|} = \frac{a_1.b_1 + a_2.b_2}{\sqrt{a_1^2 + a_2^2}.\sqrt{b_1^2 + b_2^2}}$$

– Áp dụng:

$$\begin{aligned}
cos(\vec{M}5,\vec{M}1) &= \frac{\vec{M}5.\vec{M}1}{|\vec{M}5|.|\vec{M}1|} = 0 \\
cos(\vec{M}5,\vec{M}2) &= \frac{\vec{M}5.\vec{M}2}{|\vec{M}5|.|\vec{M}2|} = 0.046 \\
cos(\vec{M}5,\vec{M}3) &= \frac{\vec{M}5.\vec{M}3}{|\vec{M}5|.|\vec{M}3|} = 0.593 \\
cos(\vec{M}5,\vec{M}4) &= \frac{\vec{M}5.\vec{M}4}{|\vec{M}5|.|\vec{M}4|} = 0
\end{aligned}$$

- Kết luận: Top 1 bộ phim khuyến nghị cho user1 là bộ phim có ID là M3.

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

- Công thức:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

- Áp dụng:

$$\begin{aligned}
J(M5,M1) &= \frac{|M5 \cap M1|}{|M5 \cup M1|} = \frac{0}{11} = 0 \\
J(M5,M2) &= \frac{|M5 \cap M2|}{|M5 \cup M2|} = \frac{1}{14-1} = 0.077 \\
J(M5,M3) &= \frac{|M5 \cap M3|}{|M5 \cup M3|} = \frac{3}{12-3} = 0.333 \\
J(M5,M4) &= \frac{|M5 \cap M4|}{|M5 \cup M4|} = \frac{0}{12} = 0
\end{aligned}$$

- Kết luận: Top 2 bộ phim khuyến nghị cho user1 là bộ phim có ID là M3 và M2.