

# HỆ KHUYẾN NGHỊ

## 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 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 User1	ID phim	Nội dung chính
	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 User1	ID phim	Nội dung chính
	M5	Tình_yêu hài_hước của cô_gái biên_kịch

Các bước tiếp theo

- Chuyển nội dung về từ thường
  - Loại bỏ từ trùng
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.

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

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.

## Biểu diễn nội dung - content(p), TF-IDF

Đưa vào một từ  $t$  và một văn bản  $d$

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) * \text{IDF}(t)$$

□ IF: Term Frequency

- Tần suất xuất hiện của từ  $t$  trong văn bản  $d$
- Những từ quan trọng xuất hiện thường xuyên hơn
- $\text{TF}(t,d) = \text{count of } t \text{ in } d / \text{number of words in } d$  (hoặc có thể không chia)

□ IDF: Inverse Document Frequency

- Giảm trọng số cho những từ xuất hiện trong hầu hết các tài liệu.

$$\text{IDF}(t) = \log \left( \frac{N}{df(t) + 1} \right)$$

- $N$ : Tổng số tài liệu
- $df(t)$ : Số tài liệu mà từ  $t$  xuất hiện trong  $N$  tài liệu

- Xây dựng bộ từ vựng từ tập dữ liệu liệu
- Tính toán TF-IDF cho từng phim
- Tính toán TF-IDF cho bộ phim User1 đã xem

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

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