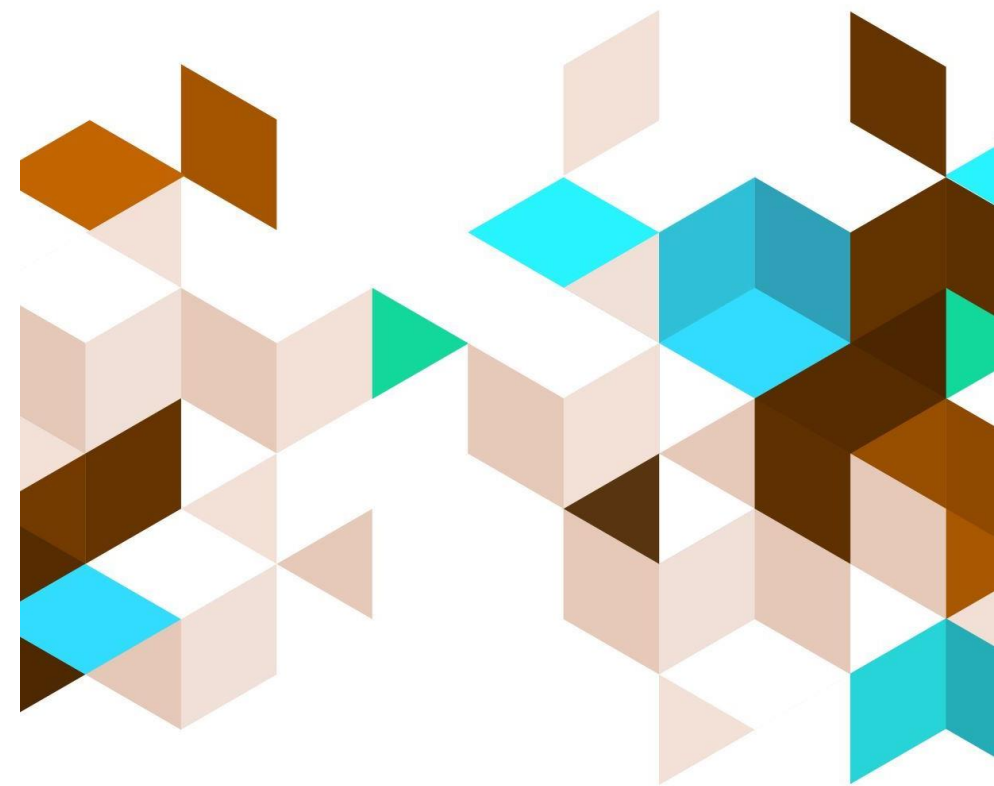


# BÁO CÁO GR2

---

TIẾP TỤC HỆ THỐNG GỢI Ý PHIM

PHẠM VIỆT HẢI 20215044



# Nội dung

---

I. Cải tiến mô hình

II. Xây dựng Web

# I. Cải tiến mô hình

---

1. Tiến trình GR1
2. Cơ sở lý thuyết
3. Demo mô hình

# 1. Tiến trình GR1

---

Neighborhood-based Collaborative Filtering

- Nhạy cảm với dữ liệu bản và thiếu
- Yêu cầu dữ liệu đầu vào lớn để có kết quả tốt

# 1. Tiến trình GR1

---

Vấn đề

	Indiana Jones	Star Wars	Empire Strikes Back	Incredibles	Casablanca
Bob	4	5	?	?	?
Ted	?	?	?	?	1
Ann	?	5	5	5	?

## 2. Cơ sở lý thuyết

---

Model-based Method: Matrix Factorization

- Dự đoán từ dữ liệu
- Tận dụng khả năng dự đoán rating của Surprise
- Kiểm thử trên kết quả gợi ý top-N

## 2. Cơ sở lý thuyết

---

Công cụ thao tác với ma trận

- PCA (Principal Component Analysis)
- SGD (Stochastic Gradient Descent)

## 2. Cơ sở lý thuyết

---

PCA:

- Giảm số chiều dữ liệu
- Giảm thiểu nhiễu
- Cải thiện hiệu suất mô hình



## 2. Cơ sở lý thuyết

PCA:

$R$

	Indiana Jones	Star Wars	Empire Strikes Back	Incredibles	Casablanca
Bob	4	5	5	4	4
Ted	3	3	3	5	4
Ann	4	5	5	5	2

PCA

	"Action"	"Sci-Fi"	"Classic"
Bob	0.3	0.5	0.2
Ted	0.1	0.1	0.8
Ann	0.3	0.6	0.1

$R^T$

	Bob	Ted	Ann
Indiana Jones	4	3	4
Star Wars	5	3	5
Empire Strikes Back	5	3	5
Incredibles	4	5	5
Casablanca	4	4	2

PCA

	"Action"	"Sci-Fi"	"Classic"
Indiana Jones	0.6	0.3	0.1
Star Wars	0.4	0.6	0
Empire Strikes Back	0.4	0.6	0
Incredibles	0.8	0.2	0
Casablanca	0.2	0	0.8

## 2. Cơ sở lý thuyết

---

SGD:

- Learning rate tùy chỉnh
- Thử trên tập mẫu ngẫu nhiên
- Tối thiểu sai số

## 2. Cơ sở lý thuyết

---

SVD và SVD++

- SVD thuần không thể chạy trên dữ liệu không đầy đủ
- SVD++ dựa trên phù hợp cho ma trận dữ liệu thiếu đa phần
- Kết quả là các ma trận phân tử hòa thiện
- Sử dụng ma trận phân tử để dự đoán rating của người dùng cho phim xác định

$$A = U\Sigma V^T$$

## 2. Cơ sở lý thuyết

---

Tuning cho SVD

```
32 param_grid = {'n_epochs': [20, 30], 'lr_all': [0.005, 0.010],
33               'n_factors': [50, 100]}
34 gs = GridSearchCV(SVD, param_grid, measures=['rmse', 'mae'], cv=3)
35
36 gs.fit(evaluationData)
37
38 # best RMSE score
39 print("Best RMSE score attained: ", gs.best_score['rmse'])
40
41 # combination of parameters that gave the best RMSE score
42 print(gs.best_params['rmse'])
```

### 3. Demo

---



# II. Xây dựng Web

---

1. Mục tiêu
2. Thiết kế
3. Xây dựng
4. Demo

# 1. Mục tiêu

---

- Thu thập dữ liệu người dùng để gợi ý
- Phát triển các chức năng trên kết quả gợi ý
- Thu thập phản hồi để cải tiến hệ thống gợi ý

## 2. Thiết kế

---

- Tổng quan
- Chức năng
- Cơ sở dữ liệu



## 2. Thiết kế

---

Chức năng

- Đánh giá
- Lưu vào profile
- Xếp hạng

## 2. Thiết kế

---

Cơ sở dữ liệu

- Người dùng
- Phim
- Thuộc tính
- Đối tượng

# 3. Xây dựng

---

- Công nghệ
- Kiến trúc: MVC
- Tiến độ

## 4. Demo

---



HẾT!

---