



**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

---

**DS300**  
**HỆ KHUYẾN NGHỊ**

# **PHƯƠNG PHÁP ĐÁNH GIÁ HỆ KHUYẾN NGHỊ**

**Giảng viên:** ThS. Nguyễn Văn Kiệt

CN. Huỳnh Văn Tín

**Bộ môn Khoa học Dữ liệu**

**Khoa Khoa học và Kỹ thuật Thông tin**

# Nội dung

---

- ❖ Giới thiệu về đánh giá
- ❖ Phương pháp Offline & Online & A/B Testing
- ❖ Một số độ đo phổ biến
  - ❖ Precision, Recall, F1
  - ❖ P@TopN, R@TopN
  - ❖ Hit Ratio, NDCG, MRR
  - ❖ #Click, #View, #CTR (Click-through rates), TimeOnRead, TimeOnPage, TimeOnSite.

# Vì sao phải đánh giá kết quả khuyến nghị

---

- Làm sao để chọn phương pháp triển khai phù hợp
- Đây là phương pháp cho kết quả tốt nhất?
- Như thế nào là tốt nhất.

# Phương pháp đánh giá kết quả khuyến nghị

---

- ❖ **Online:** Phản ánh đúng thực tế, nhưng tốn chi phí.
- ❖ **Offline:** dùng GroundTruth (Sự thật)
  - Mô phỏng quá trình online khi người dùng sử dụng kết quả khuyến nghị.
  - Lưu lại dữ liệu lịch sử, ẩn đi những tương tác của người dùng đánh giá với những thuật toán khác nhau.

# Phương pháp đánh giá kết quả khuyến nghị

---

## ❖ Xây dựng tập kiểm thử

- Chọn tập người dùng, đối tượng khuyến nghị ntn?
- Gán nhãn giá trị rating giữa người dùng và đối tượng khuyến nghị cho từng bài toán ra sao?
- Làm thế nào để có được tập kiểm thử đủ tin cậy?

## ❖ Chọn độ đo offline, online phù hợp với bài toán

# Phương pháp đánh giá kết quả khuyến nghị

---

**Đánh giá Offline**

**Không quan tâm thứ tự sắp xếp**

# MAE/NMAE

---

- ❖ **Mean Absolute Error (MAE).** Lỗi trung bình tuyệt đối giữa xếp hạng dự đoán ( $\hat{r}$ ) và xếp hạng thực của người dùng ( $r$ )

$$MAE = \frac{\sum_{ij} |\hat{r}_{ij} - r_{ij}|}{n}$$

$$NMAE = MAE / (r_{max} - r_{min})$$

# MAE/NMAE

---

❖ Ví dụ

<i>Item</i>	<i>Predicted Rating</i>	<i>True Rating</i>
1	1	3
2	2	5
3	3	3
4	4	2
5	4	1

$$MAE = \frac{|1 - 3| + |2 - 5| + |3 - 3| + |4 - 2| + |4 - 1|}{5} = 2$$

$$NMAE = \frac{MAE}{5 - 1} = 0.5$$



# RMSE/NRMSE

---

- ❖ **Root Mean Square Error (RMSE) – Lỗi trung bình bình phương gốc:** R-MSE càng nhỏ tức là sai số càng bé thì mức độ ước lượng cho thấy độ tin cậy của mô hình có thể đạt cao nhất.
- ❖ Tương tự MAE, nhưng tập trung nhiều hơn vào mức độ của độ lệch.
- ❖ Giá trị dự đoán ( $\hat{r}$ ), giá trị thực của người dùng ( $r$ )

$$RMSE = \sqrt{\frac{1}{n} \sum_{i,j} (\hat{r}_{ij} - r_{ij})^2}$$

# RMSE/NRMSE

---

❖ Ví dụ:

<i>Item</i>	<i>Predicted Rating</i>	<i>True Rating</i>
1	1	3
2	2	5
3	3	3
4	4	2
5	4	1

$$\begin{aligned} RMSE &= \sqrt{\frac{(1-3)^2 + (2-5)^2 + (3-3)^2 + (4-2)^2 + (4-1)^2}{5}} \\ &= 2.28 \end{aligned}$$

# MAE, RMSE – Vấn đề

---

MAE, RMSE được đánh giá là phù hợp cho công việc tiên đoán. Tuy nhiên nó thật sự phù hợp khi chúng ta không cần phân biệt giữa các lỗi đánh giá. Chẳng hạn, giá trị đánh giá thật sự của người dùng  $u$  trên đối tượng  $p$  là 2, nhưng kết quả hai phương pháp khác nhau dự đoán lần lượt là 1, 3. Khi đó độ lệch dùng MAE, RMSE cho kết quả như nhau. Nhưng thật ra ý nghĩa khuyến nghị sẽ khác nhau. Hay nói rõ hơn, **phương pháp dự đoán giá trị 3 sẽ ưu tiên khuyến nghị cho đối tượng  $p$  cho  $u$ , trong khi phương pháp cho giá trị 1 sẽ không mong muốn khuyến nghị cho người dùng  $u$ .**

# Độ đo Precision, Recall, F-measure

---

Bài toán khuyến nghị - phân lớp

		PREDICTED	
		Positive	Negative
ACTUAL	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}}$$

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}}$$

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

# P@TopN, R@TopN

---

- Tương tự cách tính trên Precision và Recall tuy nhiên chỉ tính trên danh sách TopN sản phẩm.
- Ví dụ: Hệ thống trả về top5 phim đề xuất cho người dùng u. Trong đó có 3 phim tồn tại trong Groundtruth của u (Groundtruth của u có 10 phim). Thì P@top5, R@top5?

$$\text{Precision} = \frac{tp}{tp + fp}$$
$$\text{Recall} = \frac{tp}{tp + fn}$$

# P@TopN, R@TopN

---

- Tương tự cách tính trên Precision và Recall tuy nhiên chỉ tính trên danh sách TopN sản phẩm.
- Ví dụ: Hệ thống trả về top5 phim đề xuất cho người dùng u. Trong đó có 3 phim tồn tại trong Groundtruth của u (Groundtruth của u có 10 phim). Thì P@top5, R@top5?

# Độ đo Precision, Recall, F-measure – vấn đề?

---

- Thứ tự trả về trong danh sách?
- Chất lượng danh sách trả về?
- Ví dụ: Hệ thống trả về top5 phim đề xuất cho người dùng u. Trong đó có 3 phim tồn tại trong Groundtruth của u (liên quan đến sở thích của u). Thì  $P@top5$ ? Nếu thuật toán A trả về 3 phim này nằm đầu danh sách và thuật toán B nằm cuối danh sách ☐ nên chọn thuật toán A, hay thuật toán B, vì sao?

# Phương pháp đánh giá kết quả khuyến nghị

---

**Đánh giá Offline**

**Quan tâm thứ tự sắp xếp**



# MRR: Mean Reciprocal Rank

---

MRR: Xếp hạng đối ứng trung bình

$$MRR = \frac{1}{|U|} \sum_{u \in U} \frac{1}{rank_u}$$

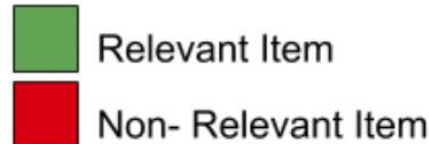
U: tập người dùng

u: u thuộc U

$rank_u$  biểu thị thứ hạng của đối tượng có liên quan đầu tiên.

# MRR: Mean Reciprocal Rank

Ví dụ



$$MRR = \frac{1}{|U|} \sum_{u \in U} \frac{1}{rank_u}$$

		Reciprocal Rank
User 1		1/3
User 2		1/2
User 3		1
Mean Reciprocal Rank	$(1/3 + 1/2 + 1) / 3 = 0.61$	

# Average Precision và Mean Average Precision

---

**Average Precision:** Chỉ số đánh giá xem tất cả các mục có liên quan đến groundtruth mà mô hình chọn có được xếp hạng cao hơn hay không. (trên 1 user)

$$AP = \frac{\sum_{k=1}^n (P(k) * rel(k))}{\text{number of relevant items}}$$

- $rel(k)$  là một hàm chỉ bằng 1 khi mục ở vị trí  $k$  có liên quan.
- $P(k)$  là chỉ số Precision @  $k$

# Average Precision và Mean Average Precision

## Average Precision – Ví dụ

$$AP = \frac{\sum_{k=1}^n (P(k) * rel(k))}{\text{number of relevant items}}$$

Thuật toán A



Precision@K    1    1/2    2/3    2/4    3/5

$$AP = \frac{(1 + 2/3 + 3/5)}{3} = 0.7555$$

Thuật toán B



Precision@K    1    1    1    3/4    3/5

$$AP = \frac{(1 + 1 + 1)}{3} = 1$$

# Average Precision và Mean Average Precision

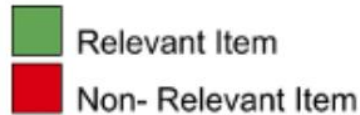
---

**Mean Average Precision : Nếu chúng ta muốn đánh giá Average Precision trên nhiều danh sách nhiều người dùng**

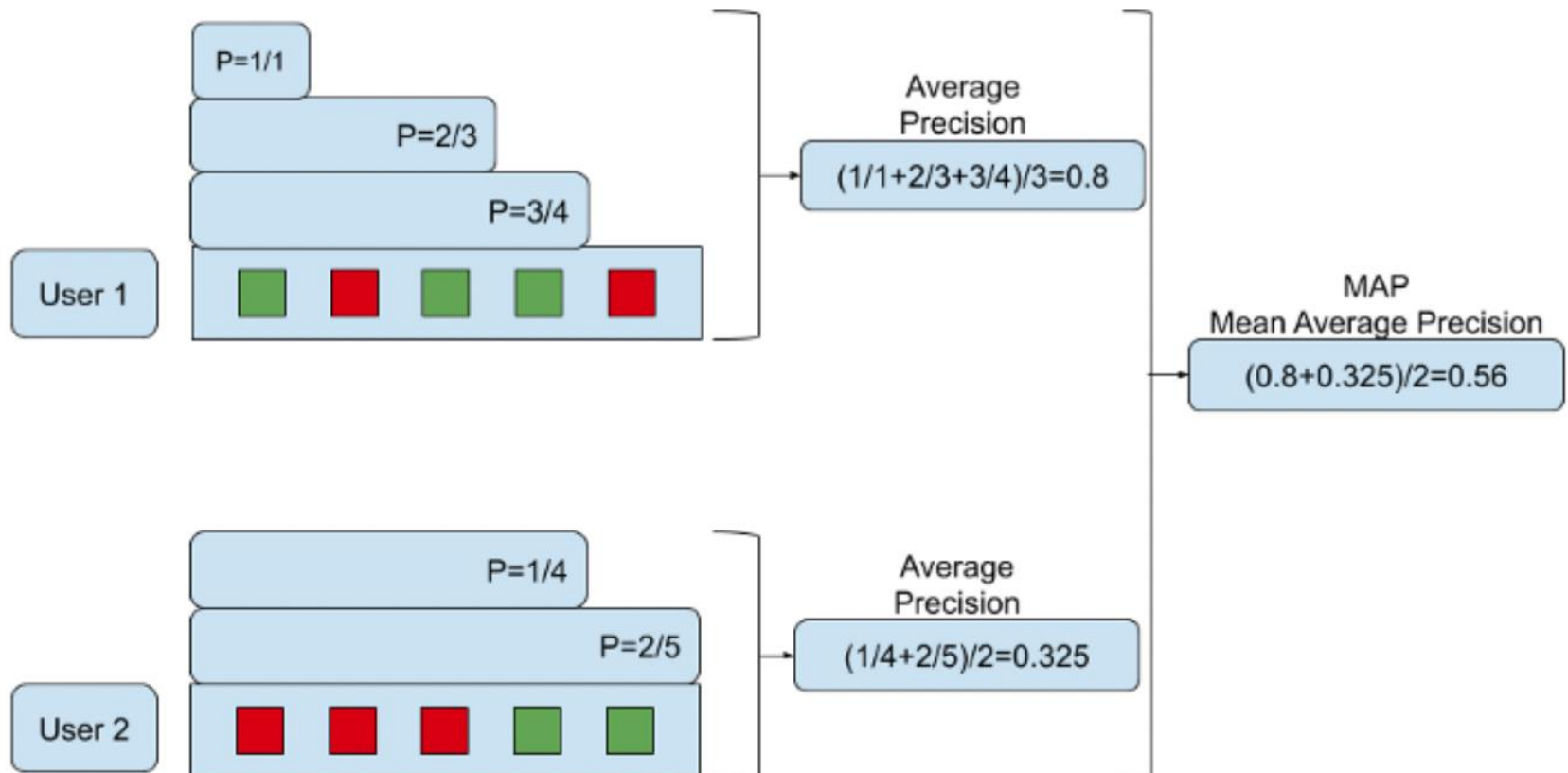
$$MAP = \frac{1}{U} \sum_{u=1}^U AP(u)$$

- U là tổng số người dùng
- AP(u) là độ chính xác trung bình trên người dùng u

# Average Precision và Mean Average Precision



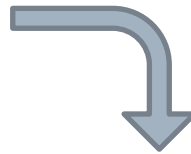
## Mean Average Precision – Ví dụ



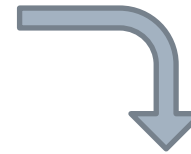
# Normalized Discounted Cumulative Gain (NDCG)

---

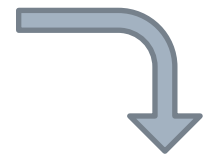
**1 - Cumulative Gain (CD@K)**



**2 - Discounted Cumulative Gain (DCG@k)**



**3 - Ideal Discounted Cumulative Gain (IDCG@k)**



**4 - Normalized Discounted Cumulative Gain (NDCG@k)**

# Normalized Discounted Cumulative Gain (NDCG)

## 1 - Cumulative Gain (CD@K)

$$CG@k = \sum_{i=1}^k rel_i$$

Relevance  
Position



Position(k)	1	2	3	4	5
Cumulative Gain@k	3	3+2=5	3+2+3=8	3+2+3+0=8	3+2+3+0+1=9



# Normalized Discounted Cumulative Gain (NDCG)

## 2 - Discounted Cumulative Gain (DCG@k)

$$DCG@k = \sum_{i=1}^k \frac{rel_i}{\log_2(i+1)}$$

Position 1 2

Model 1 3 0 CG@2 = 3+0 = 3

Model 2 0 3 CG@2 = 0+3 = 3

↑ same

Relevance	3	2	3	0	1
Position	1	2	3	4	5

Position(i)	Relevance(rel <sub>i</sub> )	log <sub>2</sub> (i + 1)	$\frac{rel_i}{\log_2(i+1)}$
1	3	log <sub>2</sub> (1 + 1) = log <sub>2</sub> (2) = 1	3 / 1 = 3
2	2	log <sub>2</sub> (2 + 1) = log <sub>2</sub> (3) = 1.5849625007211563	2 / 1.5849 = 1.2618
3	3	log <sub>2</sub> (3 + 1) = log <sub>2</sub> (4) = 2	3 / 2 = 1.5
4	0	log <sub>2</sub> (4 + 1) = log <sub>2</sub> (5) = 2.321928094887362	0 / 2.3219 = 0
5	1	log <sub>2</sub> (5 + 1) = log <sub>2</sub> (6) = 2.584962500721156	1 / 2.5849 = 0.3868

# Normalized Discounted Cumulative Gain (NDCG)

## 3 - Ideal Discounted Cumulative Gain (DCG@k)

$$DCG@k = \sum_{i=1}^k \frac{rel_i}{\log_2(i+1)}$$



Let's calculate the ideal DCG(IDC@k) for this order.

Position( <i>i</i> )	Relevance( <i>rel<sub>i</sub></i> )	$\log_2(i+1)$	$\frac{rel_i}{\log_2(i+1)}$	IDCG@k
1	3	$\log_2(2) = 1$	$3 / 1 = 3$	3
2	3	$\log_2(3) = 1.5849$	$3 / 1.5849 = 1.8927$	$3 + 1.8927 = 4.8927$
3	2	$\log_2(4) = 2$	$2 / 2 = 1$	$3 + 1.8927 + 1 = 5.8927$
4	1	$\log_2(5) = 2.3219$	$1 / 2.3219 = 0.4306$	$3 + 1.8927 + 1 + 0.4306 = 6.3233$
5	0	$\log_2(6) = 2.5849$	$0 / 2.5849 = 0$	$3 + 1.8927 + 1 + 0.4306 + 0 = 6.3233$

# Normalized Discounted Cumulative Gain (NDCG)

---

## 4 - Normalized Discounted Cumulative Gain (NDCG@k)

$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

<i>k</i>	DCG@k	IDCG@k	NDCG@k
1	3	3	3 / 3 = 1
2	4.2618	4.8927	4.2618 / 4.8927 = 0.8710
3	5.7618	5.8927	5.7618 / 5.8927 = 0.9777
4	5.7618	6.3233	5.7618 / 6.3233 = 0.9112
5	6.1486	6.3233	6.1486 / 6.3233 = 0.9723

# Phương pháp đánh giá kết quả khuyến nghị

---

**Đánh giá Online**

# Độ đo cho đánh giá online

---

- Số lượng click: #click
- Số lượng view: #view
- CTR: Click-through rates

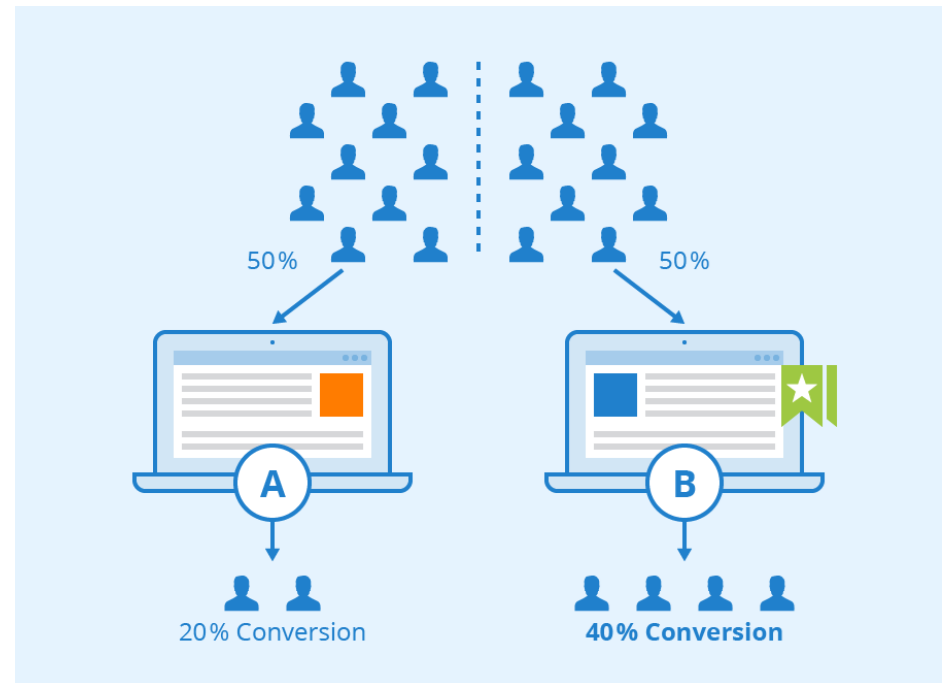
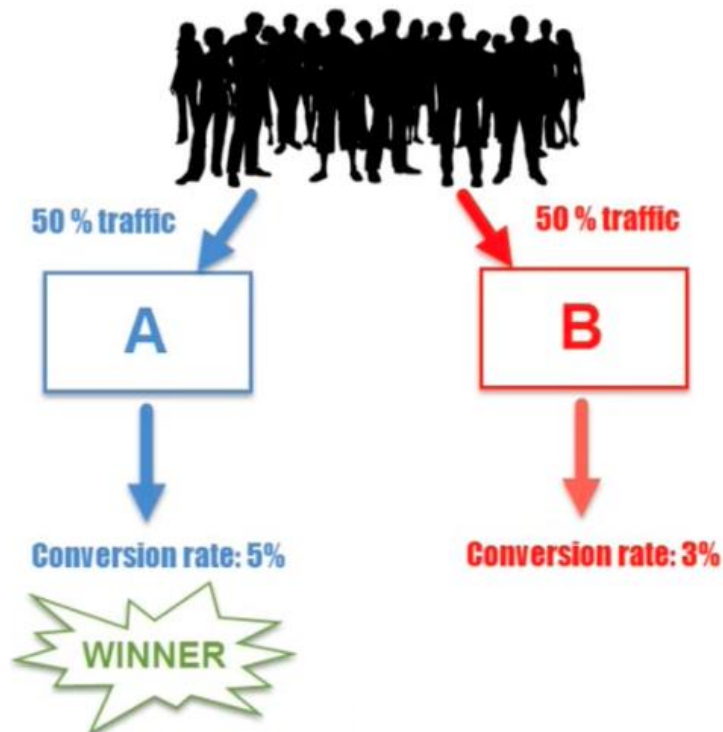
# Độ đo cho đánh giá online

---

- TimeOnRead: #Thời gian đọc
- TimeOnPage: Thời gian ở lại trang
- TimeOnSite: Thời gian ở lại site

# A/B Testing

Nhằm phân tích và hiểu rõ hành vi, phản hồi của người dùng



# Q&A



**Thank you!**