



ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC BÁCH KHOA

BÀI GIẢNG HỆ THỐNG THÔNG TIN

D
BACH KHOA
N
A
N
G
P



Khoa Công Nghệ Thông Tin
TS. Nguyễn Văn Hiệu

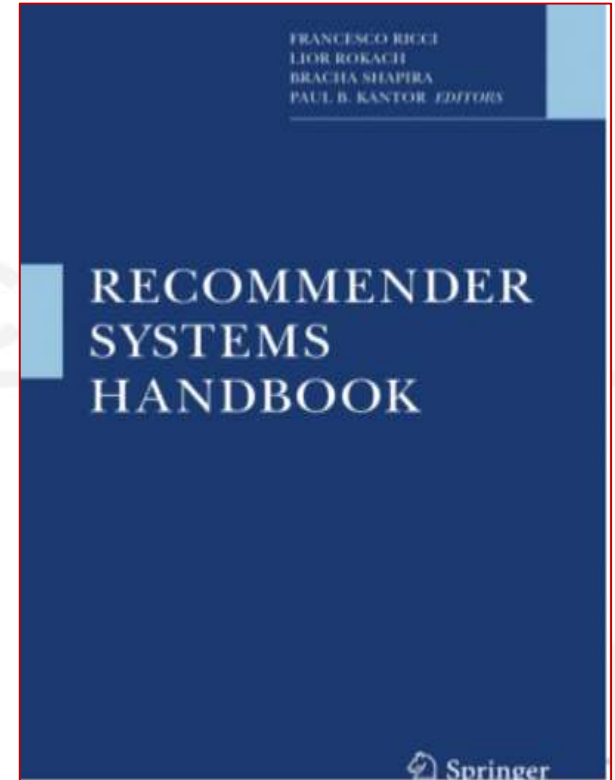
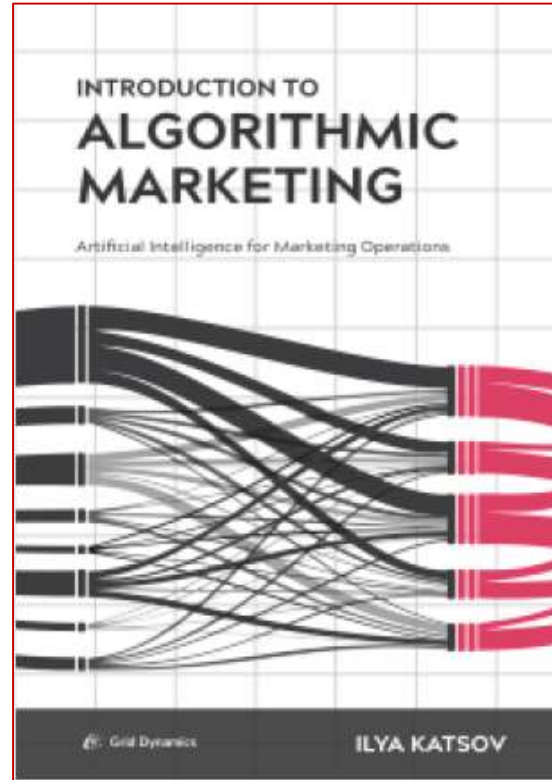
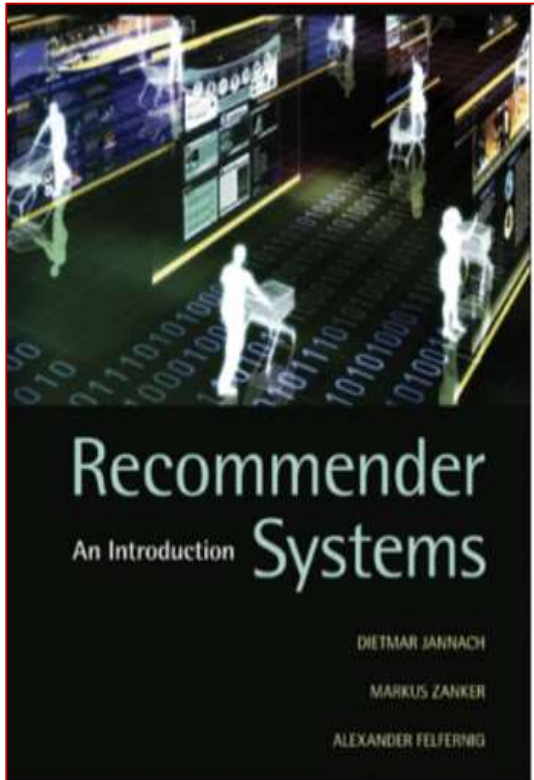
BÀI GIẢNG HỆ THỐNG THÔNG TIN VÀ ERP

Bài 5: HTTT RS

Nội dung

- Giới thiệu
- Mục đích RS
- Biểu diễn dữ liệu
- Bài toán RS
- Phân loại phương pháp
- Chi tiết vài phương pháp
- Trao đổi

Tài liệu



- Vấn đề: “Information overload”
 - Trồng cây gì? Nuôi con gì?
 - Đọc sách nào? Xem phim gì? Mua Tablet nào?
 - Đi du lịch ở đâu ? Chọn tour nào?
 - Đi cùng ai?

Giới thiệu

Đọc sách nào ?



Xem phim gì ?





Giới thiệu

Mua tablet nào ?



Giải pháp:

+ Cần **decision support system**

=> Cần **recommendation system**

Giới thiệu

- Khởi nguồn cảm hứng
 - Dữ liệu: 100 M rating (1-5 stars) from 480K users on 18K movies.

Netflix Prize

Home Rules Leaderboard Register Update Submit Download

Leaderboard 10.05% Display top 20 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8558	10.05	2009-06-26 18:42:37
Grand Prize - RMSE <= 0.8563				
2	Pragmatic Theory	0.8582	9.80	2009-06-25 22:15:51
3	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
4	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
5	Dace	0.8604	9.56	2009-04-22 05:57:03
6	BigChaos	0.8613	9.47	2009-06-23 23:06:52



Giới thiệu

- Dự đoán Bob “**thích**” dâu tây ?



Nguồn: Lester Mackey, 2009

- Giả định: Sở thích là “đồng liên quan” dự đoán khả năng anh thích đến Z ?
 - Nếu em thích “**đến**” X, Y và Z
 - Nếu anh thích “**đến**” X và Y

100% 😊

Giới thiệu

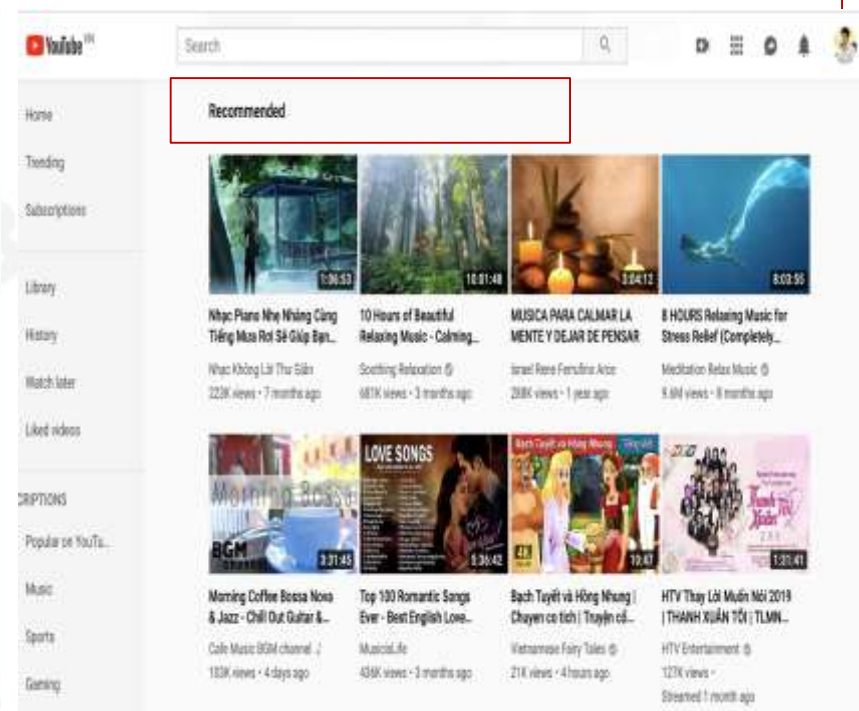
- “Nổi tiếng”: gợi ý bán hàng trực tuyến/ comment

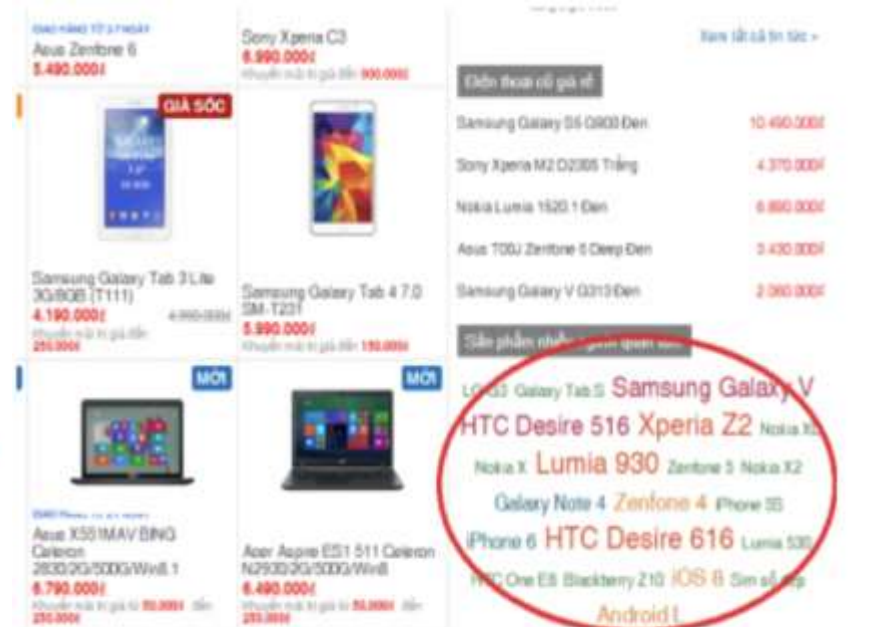
The screenshot displays an Amazon product page for the "ArmorSuit MilitaryShield - Samsung Galaxy S3 Screen Protector". The page is divided into several sections:

- Top Navigation:** Includes links for "Your Amazon.com", "Your Browser History", "Recommended For You", "Amazon Betterizer", "Improve Your Recommendations", and "Your Profile".
- Recommended for You:** A section highlighted with a blue circle, showing "These recommendations are based on items you own and more." with options to view "All", "New Releases", or "Coming Soon".
- Just For Today:** A section with a "Browse Recommended" link.
- Recommendations:** A list of links for various categories: Amazon Instant Video, Amazon MP3 Store, Appliances, Appstore for Android, Arts, Crafts & Sewing, Automotive, Baby, Beauty, Books, Books on Kindle, Camera & Photo, and Cell Phones & Accessories.
- Product Listing:**
 - Item 1:** "ArmorSuit MilitaryShield - Samsung Galaxy S3 Screen Protector" by ArmorSuit (May 18, 2012). It has an average customer review of 4.5 stars (344 reviews), is in stock, and has a list price of \$42.95 with a current price of \$9.95. It is marked as "6 used & new from \$9.17".
 - Item 2:** "Aduro Shell Holster Combo Case for Samsung Galaxy S3" by Aduro (June 13, 2012). It has an average customer review of 4.5 stars (113 reviews) and is in stock as of September 27, 2012.
- Customer Reviews:**
 - Most Helpful Customer Reviews:** A section highlighted with a red circle, showing two reviews. The first review is by Raj Rameshpriya on August 10, 2012, titled "Amazing introduction to NoSQL, and scalability". The second review is by Mark on August 10, 2012, titled "A Must Read".
 - Most Recent Customer Reviews:** A section highlighted with a red circle, showing a review by "Extremely well written book on NoSQL concepts" dated August 10, 2012.
- Customer Images:** A section with an "Add an image" button.

Giới thiệu

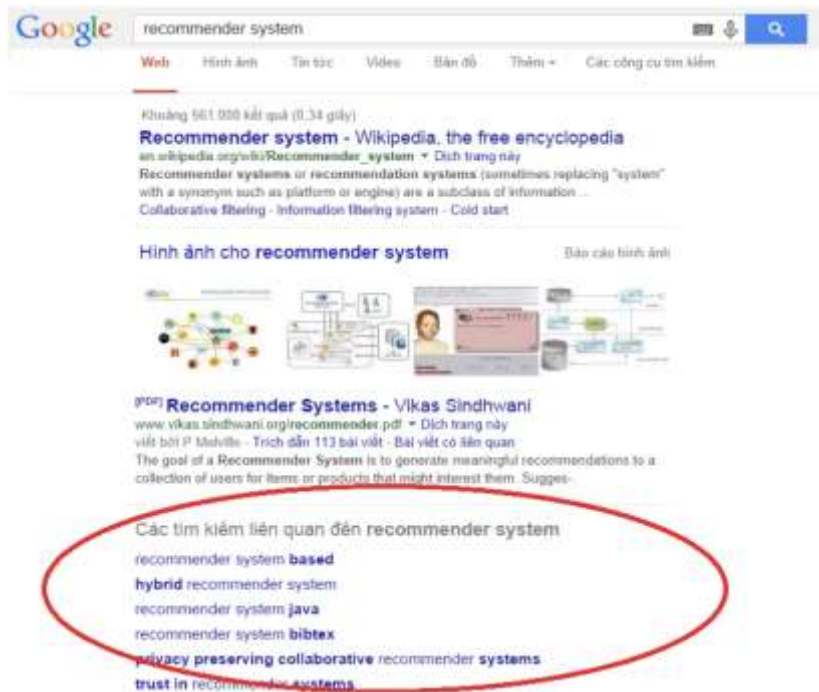
- "Nổi tiếng": gợi ý giải trí (music, video, movie)



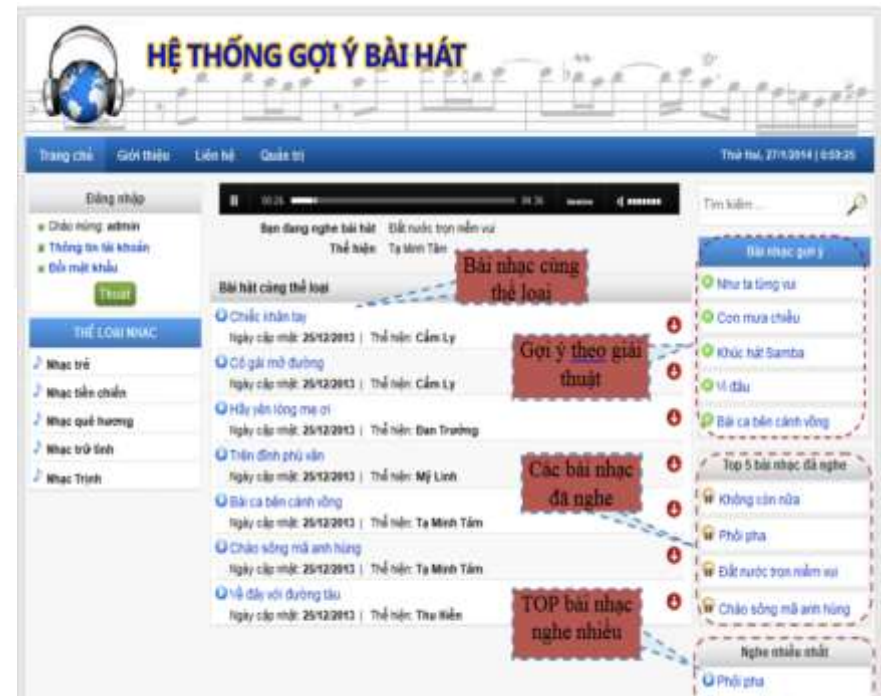


Giới thiệu

Gợi ý theo keyword



Gợi ý theo new/ populator item



Mục đích RS

- Mục đích RS là dựa vào “**sở thích**” (*thói quen/ nhu cầu/ năng lực, /...*) trong quá khứ của user để dự đoán sở thích trong tương lai, từ đó đưa ra **danh sách gợi ý tương ứng**
- Tùy theo hệ thống, feedback từ user (*để ước lượng mức độ thích*) có thể khác nhau:



○ Feedback “rõ ràng”:

- + số 1, 2,..., 5.
- + like (1) / dislike(0)
- +

○ Feedback “ngầm”:

- + thời gian quan sát trên mỗi item
- + số lần kích chuột,
- +

Dữ liệu Feedback === > kỹ thuật gợi ý

Biểu diễn dữ liệu cho RS

	<i>Items</i>					
	<i>1</i>	<i>2</i>	<i>...</i>	<i>i</i>	<i>...</i>	<i>m</i>
<i>1</i>	5	3		1	2	
<i>2</i>		2				4
<i>:</i>			5			
<i>u</i>	3	4	?	2	1	
<i>:</i>					4	
<i>n</i>			3	2		

- $\hat{r} : U \times I \rightarrow R$
- \hat{r}_{ui} : xếp hạng của người dùng u cho sản phẩm i
- Dự đoán giá trị cho các ô trống

Mô tả bài toán RS

- Gọi

- **U**: user IDs

- **I**: item IDs

- **R** : rating

- Tập dữ liệu $D: U \times I \times R$

- $D^{Train} \subseteq D$

- $D^{Test} \subseteq D$

- Bài toán: cho D^{Train} , Tìm $\hat{r} : U \times I \rightarrow R$:

- $\varepsilon(\hat{r}, r)$ thỏa mãn điều kiện cho trước với $(u, i, r) \in D^{Test}$.

- Ví dụ: nếu ε là RMSE (root mean squared error) thì $\varepsilon(\hat{r}, r)$ cần phải tối thiểu.

$$RMSE = \sqrt{\frac{\sum_{(u,i,r) \in D^{test}} (r - \hat{r}_{(u,i)})^2}{|D^{test}|}}$$

Mô tả bài toán RS

Dữ liệu ví dụ

Training data

user	Item	rating
1	21	1
1	213	5
2	345	4
2	123	4
2	768	3
3	76	5
4	45	4
5	568	1
5	342	2
5	234	2
6	76	5
6	56	4

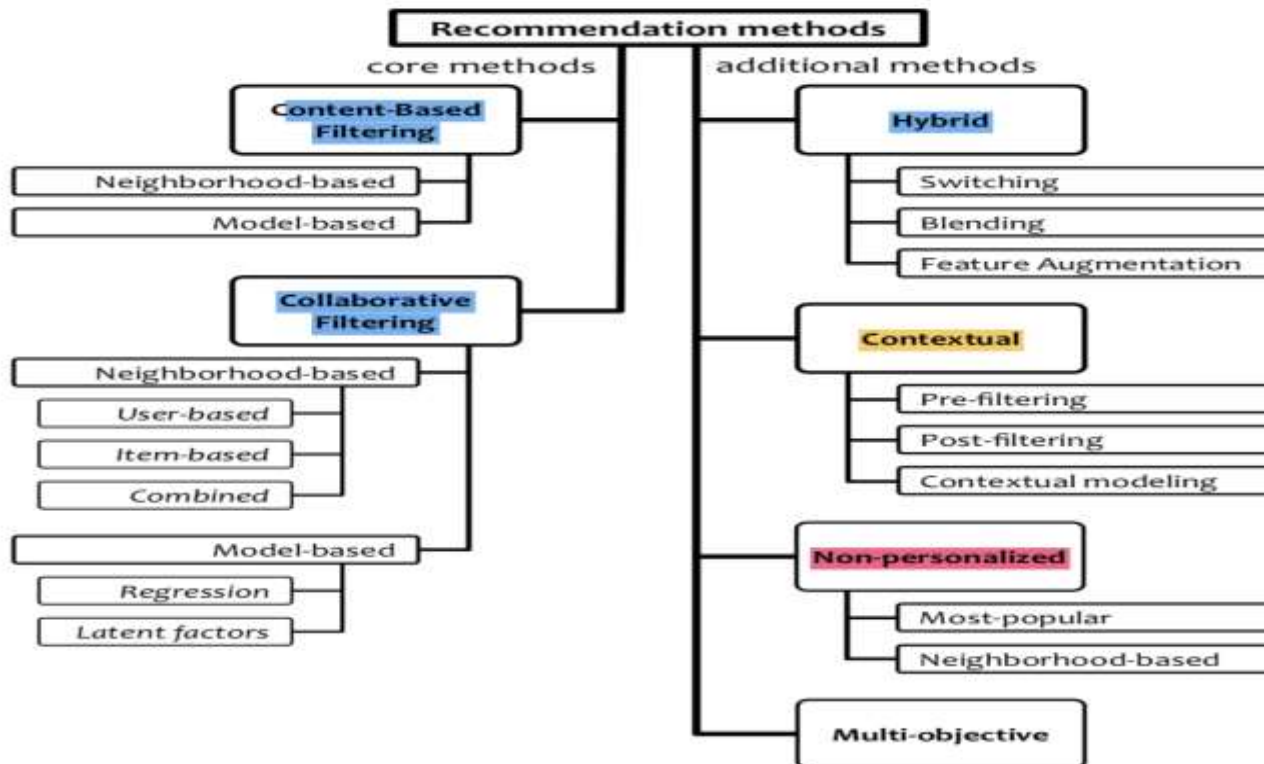
Test data

user	Item	rating
1	62	?
1	96	?
2	7	?
2	3	?
3	47	?
3	15	?
4	41	?
4	28	?
5	93	?
5	74	?
6	69	?
6	83	?

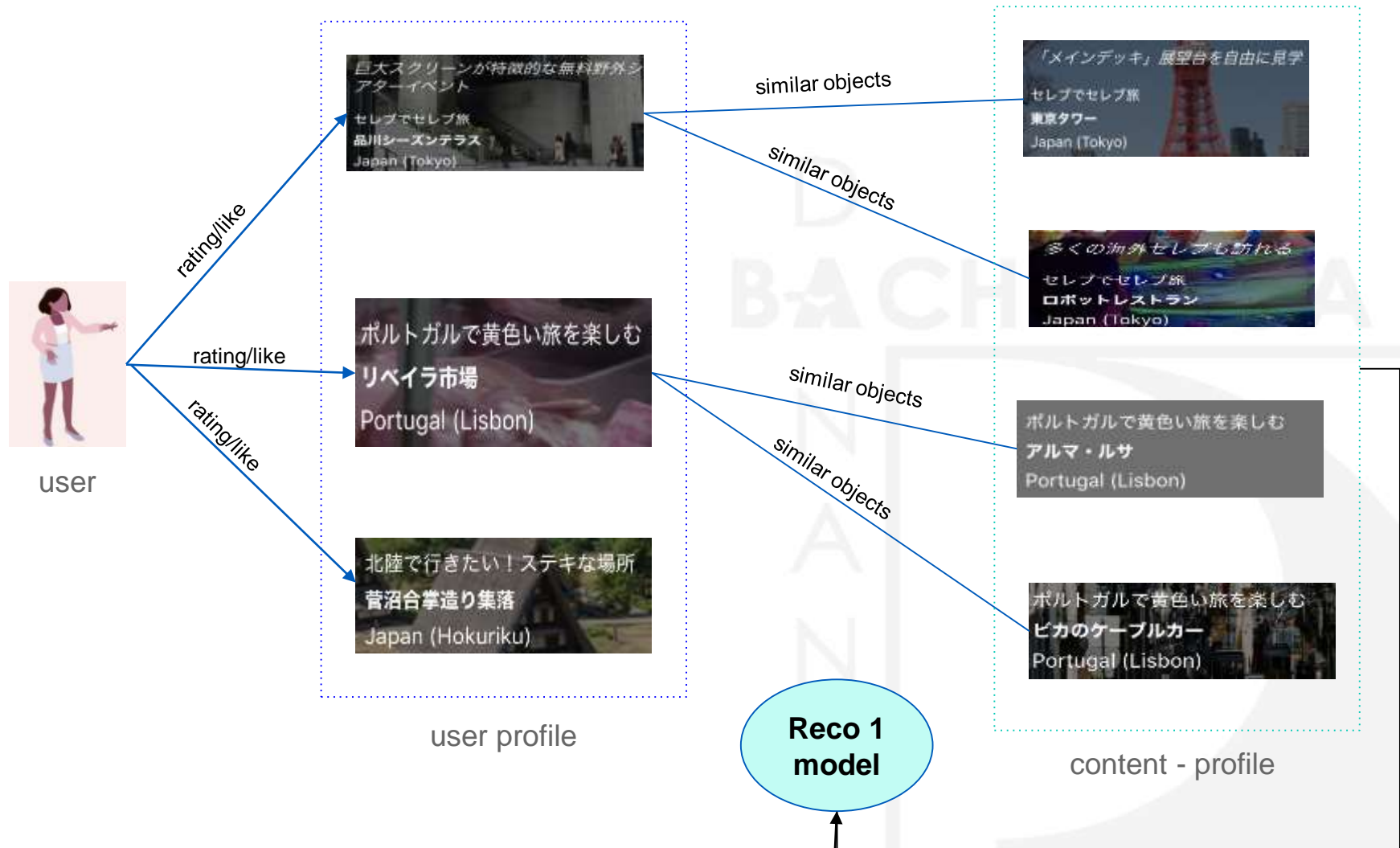
Nhóm kỹ thuật gợi ý chính



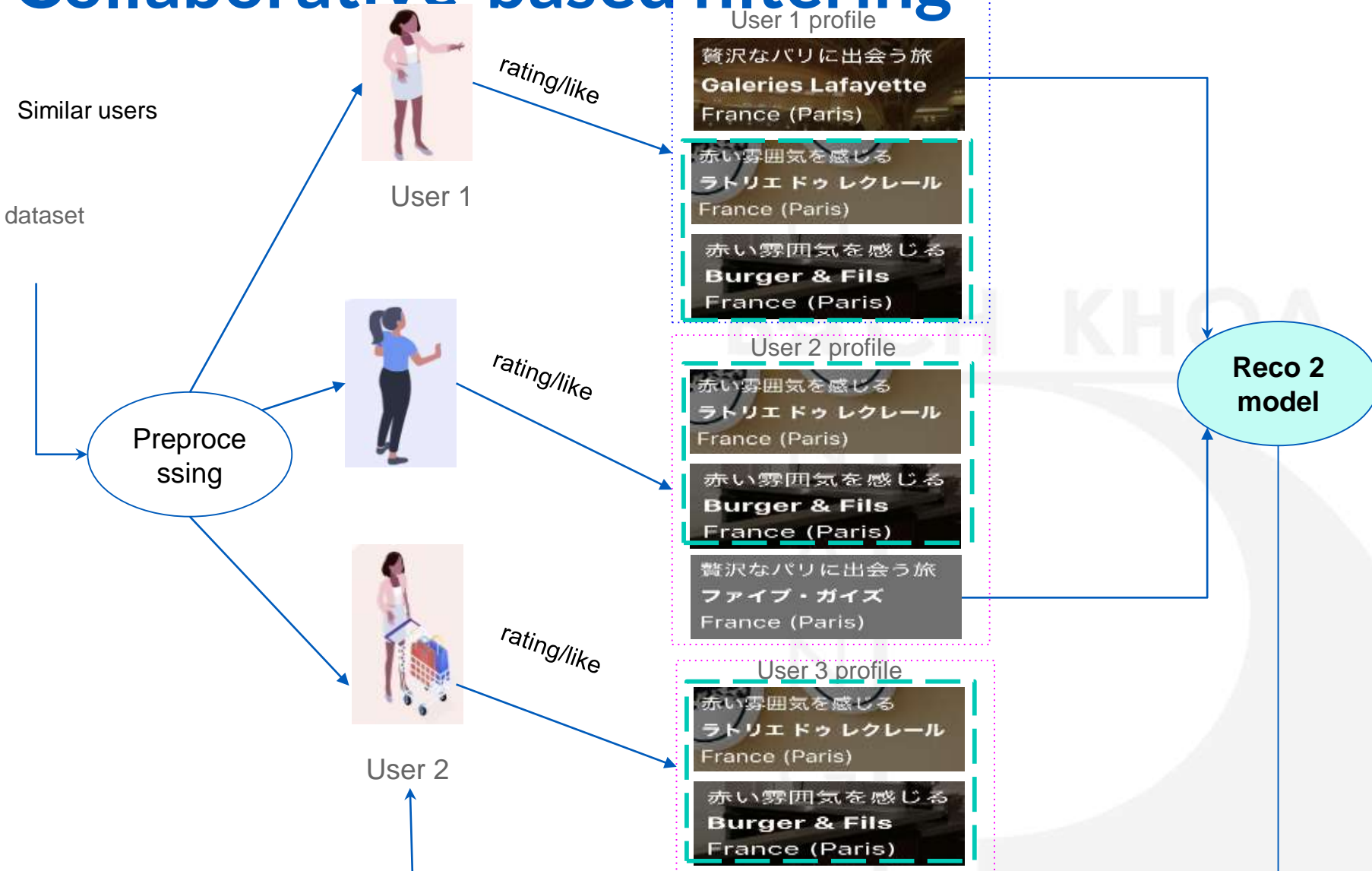
Nhóm kỹ thuật gợi ý chính



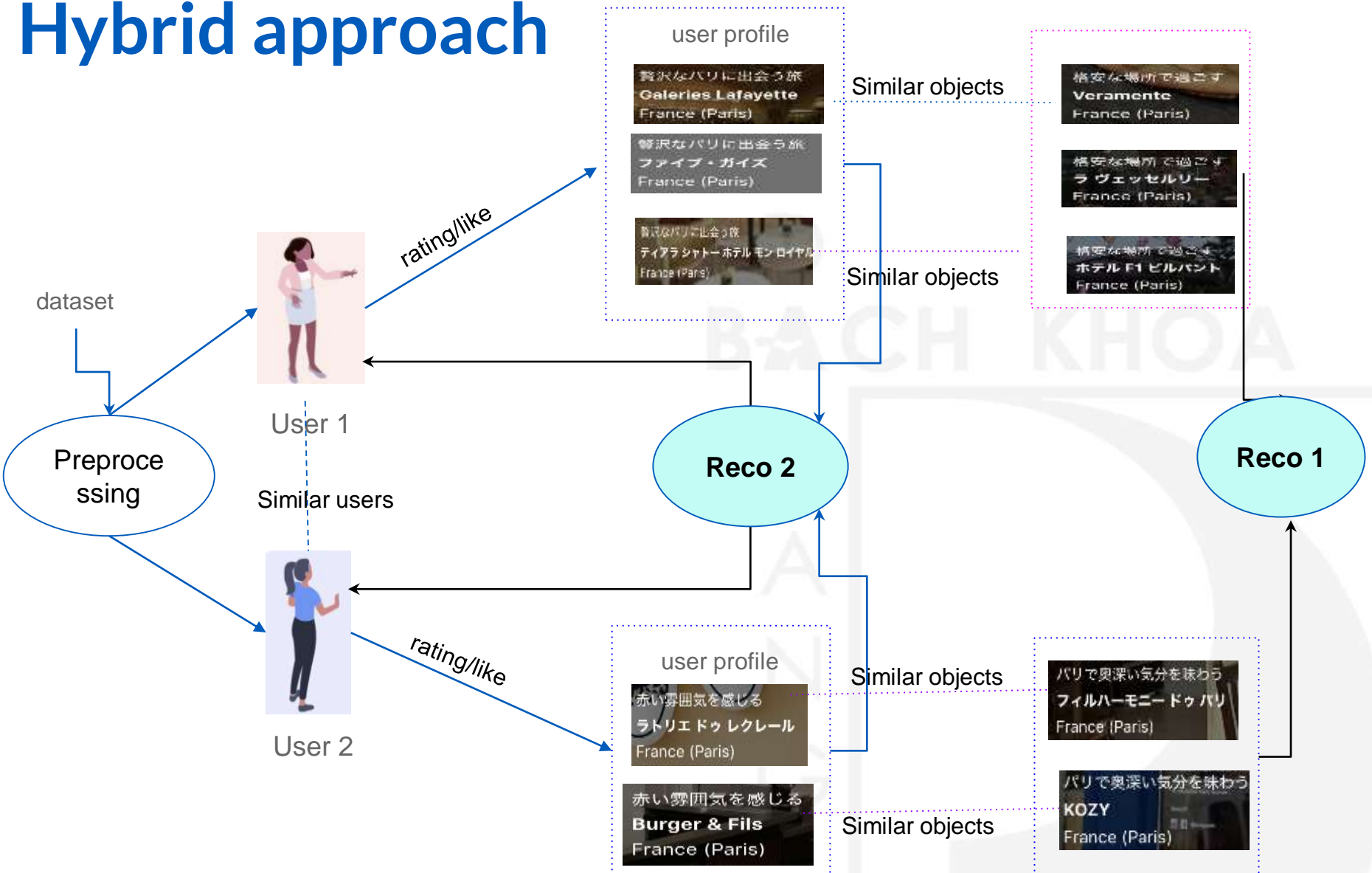
Content-based filtering



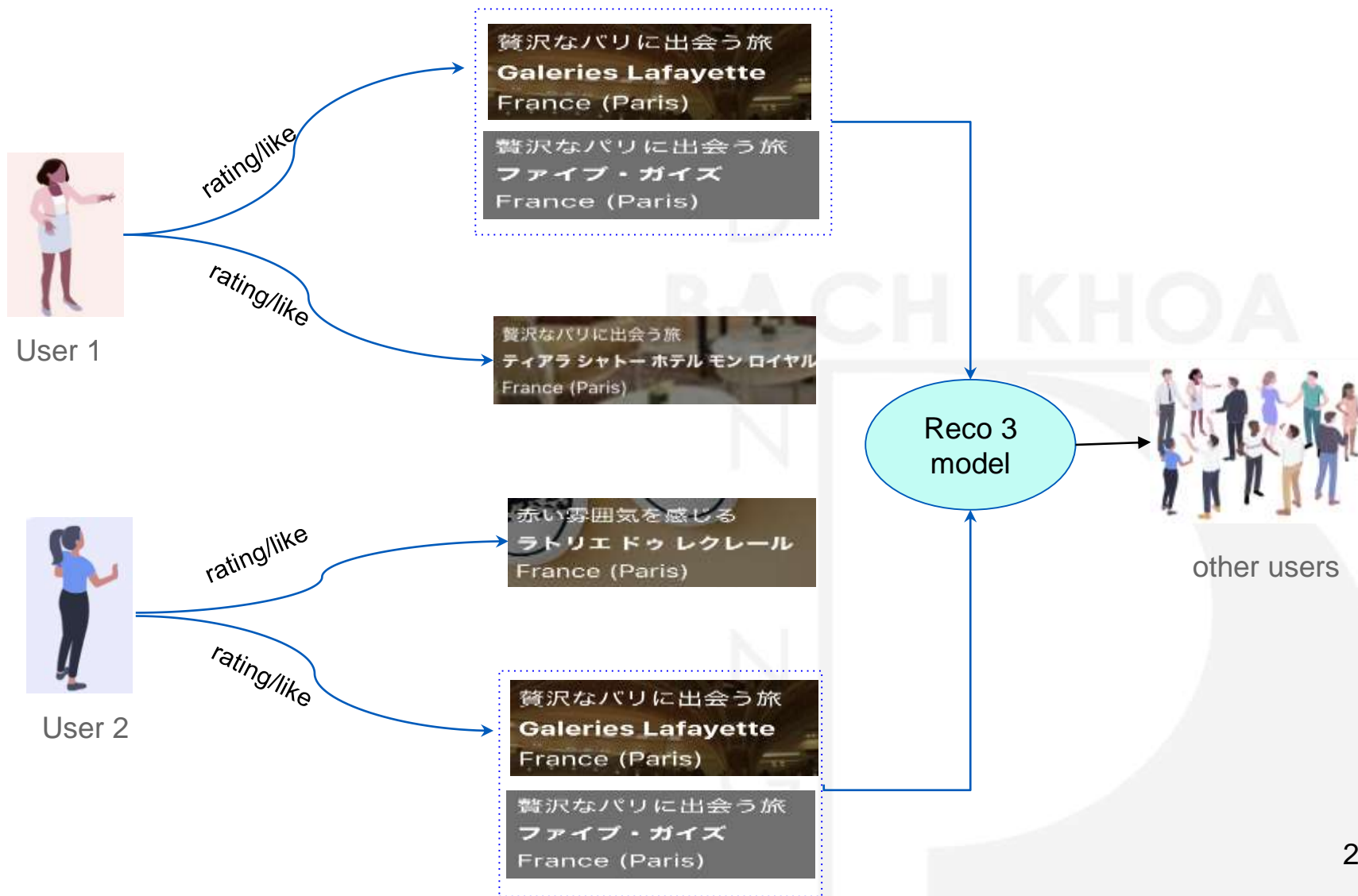
Collaborative-based filtering



Hybrid approach

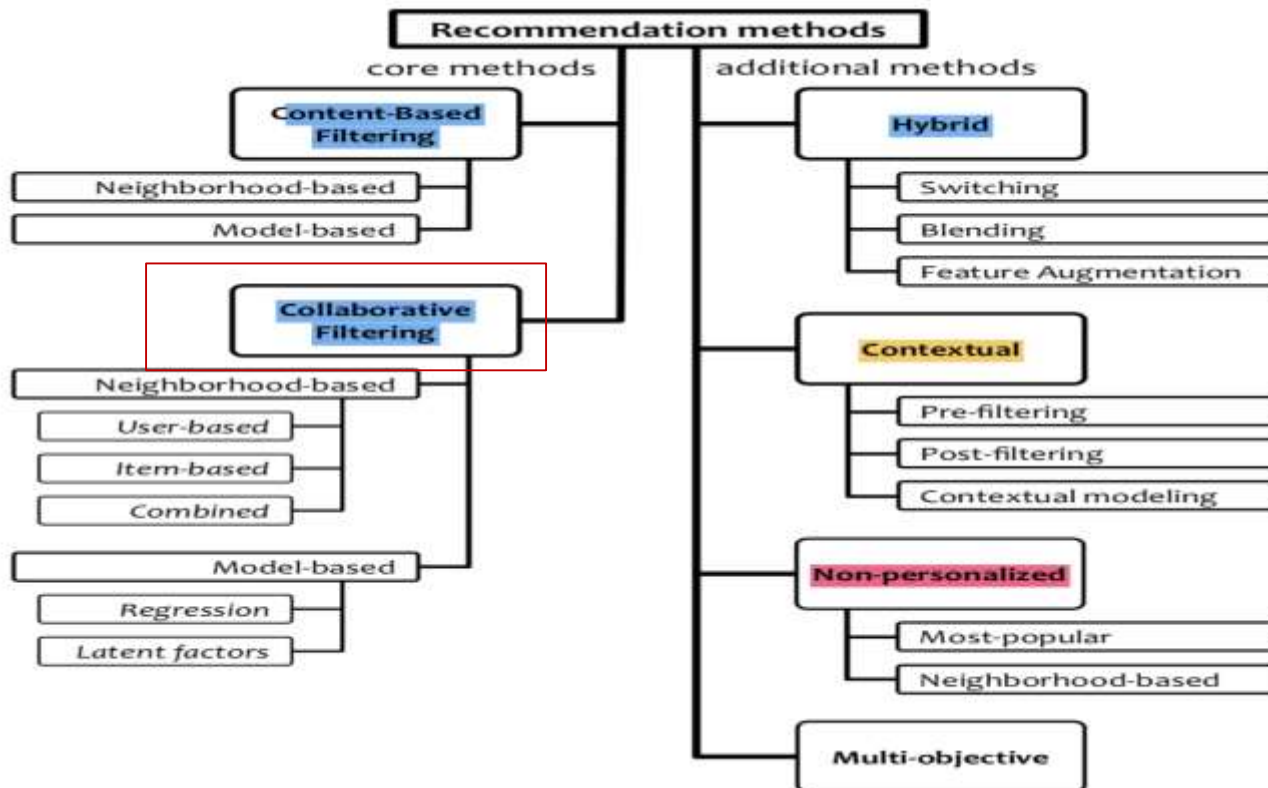


Popular-based approach



Association rule approach





Lọc cộng tác (Collaborative filtering)

- Cộng tác = sử dụng dữ liệu của người khác
- kỹ thuật láng giềng (Neighborhood-based)
 - Cơ sở người dùng (user -base)
 - Cơ sở sản phẩm(item- base)
- kỹ thuật dự vào mô hình (model based)
 - Talent factor model
 - **Matrix factorization model**

Kỹ thuật láng giềng

- Vấn đề: Đánh giá của người dùng (user) với sản phẩm (item)

Giải pháp 1: User-base

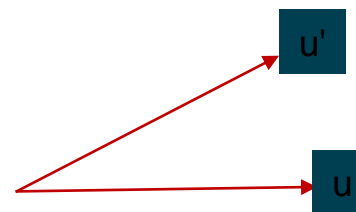
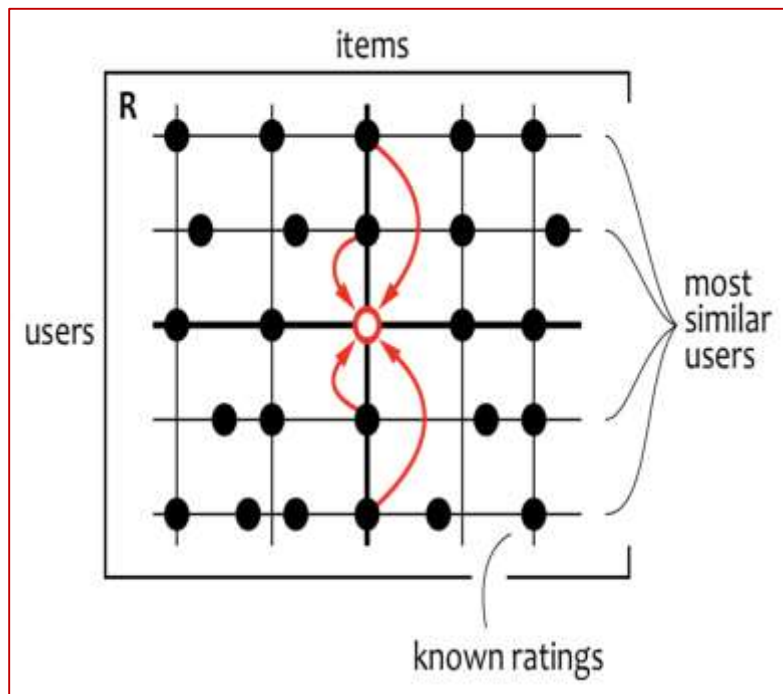


Giải pháp 2: Item-base



Kỹ thuật láng giềng

- Giải pháp 1: user-base

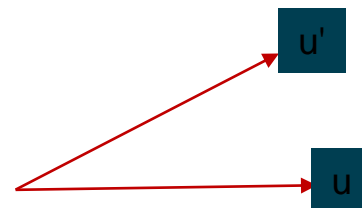
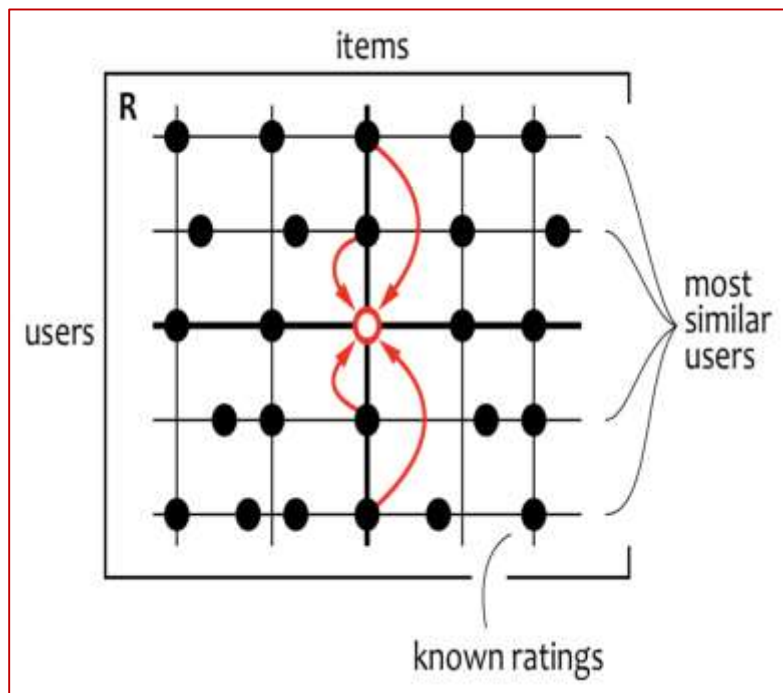


$$\text{sim}_{\text{cosine}}(u, u') = \frac{\sum_{i \in I_{uu'}} r_{ui} \cdot r_{u'i}}{\sqrt{\sum_{i \in I_{uu'}} r_{ui}^2} \sqrt{\sum_{i \in I_{uu'}} r_{u'i}^2}}$$

$$\hat{r}_{ui} = \frac{\sum_{u' \in K_u} \text{sim}(u, u') \cdot r_{u'i}}{\sum_{u' \in K_u} |\text{sim}(u, u')|}$$

Kỹ thuật láng giềng

- Giải pháp 1: user-base



$$\text{sim}_{\text{pearson}}(u, u') = \frac{\sum_{i \in I_{uu'}} (r_{ui} - \bar{r}_u)(r_{u'i} - \bar{r}_{u'})}{\sqrt{\sum_{i \in I_{uu'}} (r_{ui} - \bar{r}_u)^2 \sum_{i \in I_{uu'}} (r_{u'i} - \bar{r}_{u'})^2}}$$

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{u' \in K_u} \text{sim}(u, u') \cdot (r_{u'i} - \bar{r}_{u'})}{\sum_{u' \in K_u} |\text{sim}(u, u')|}$$

Kỹ thuật láng giềng

Recommendation tasks: Example

Rating prediction from explicit feedback

- How would Steve rate the Titanic movie?

	Titanic	Pulp Fiction	Iron Man	Forrest Gump	The Mummy
Joe	1	4	5		3
Ann	5	1		5	2
Mary	4	1	2	5	
Steve	?	3	4		4

Item recommendation from implicit feedback

- Which movie(s) Steve would like to see/buy?

	Titanic	Pulp Fiction	Iron Man	Forrest Gump	The Mummy
Joe	1	1	1		1
Ann	1	1		1	1
Mary	1	1	1	1	
Steve	?	1	1	?	1

Thank Tomas Horváth for this example!

kỹ thuật láng giềng

User similarity: Example

Cosine similarity:

$sim(u, u')$	Joe	Ann	Mary	Steve
Joe	1.0	0.283	0.372	0.962
Ann	—	1.0	0.915	0.232
Mary	—	—	1.0	0.254
Steve	—	—	—	1.0

Pearson similarity:

$sim(u, u')$	Joe	Ann	Mary	Steve
Joe	1.0	-0.716	-0.762	-0.005
Ann	—	1.0	0.972	0.565
Mary	—	—	1.0	0.6
Steve	—	—	—	1.0

Kỹ thuật láng giềng

Prediction using 2 most similar users: Example

rating prediction using 2 most similar users:

$$\triangleright U_{Titanic} = \{Joe, Ann, Mary\},$$

$$K_{Steve,2}^{Titanic} = \{Mary, Ann\}$$

$$\triangleright \bar{r}_{Steve} = \frac{11}{3} = 3.67 \quad \bar{r}_{Mary} = \frac{12}{4} = 3 \quad \bar{r}_{Ann} = \frac{13}{4} = 3.25$$

Using Pearson sim:

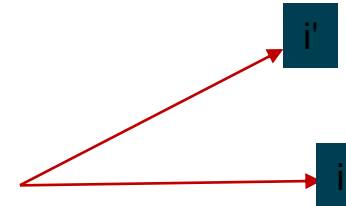
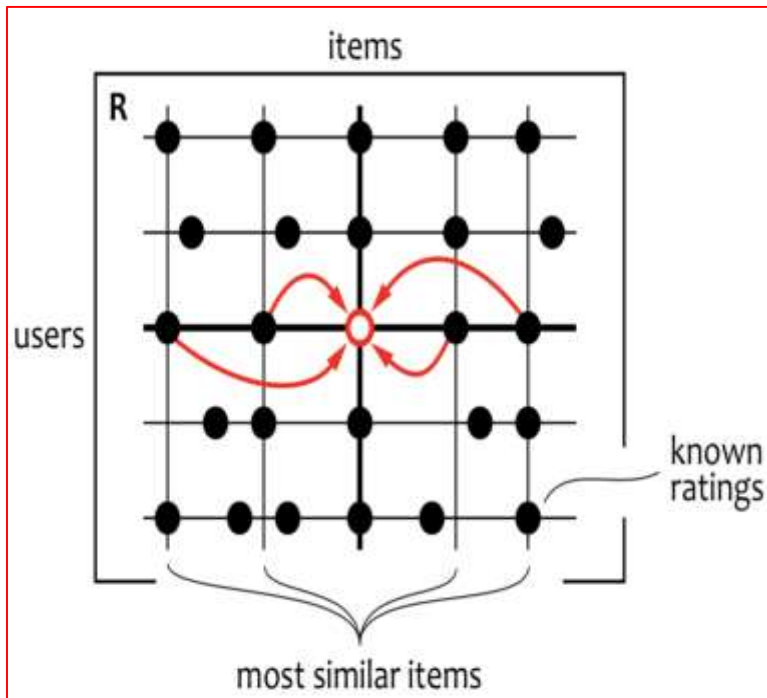
$$\triangleright \hat{r}_{ST} = \bar{r}_S + \frac{sim(S,M) \cdot (r_{MT} - \bar{r}_M) + sim(S,A) \cdot (r_{AT} - \bar{r}_A)}{|sim(S,M)| + |sim(S,A)|} =$$

$$3.67 + \frac{0.6 \cdot (4 - 3) + 0.565 \cdot (5 - 3.25)}{0.6 + 0.565} = 1.36$$

$$\hat{r}_{ui} = \bar{r}_i + \frac{\sum_{i' \in K_i} \text{sim}(i, i') \cdot (r_{ui'} - \bar{r}_{i'})}{\sum_{i' \in K_i} |\text{sim}(i, i')|}$$

Kỹ thuật láng giềng

- Giải pháp 2: item-base



$$sim_{pearson}(i, i') = \frac{\sum_{u \in U_{i, i'}} (r_{ui} - \bar{r}_i)(r_{ui'} - \bar{r}_{i'})}{\sqrt{\sum_{u \in U_{i, i'}} (r_{ui} - \bar{r}_i)^2} \cdot \sqrt{\sum_{u \in U_{i, i'}} (r_{ui'} - \bar{r}_{i'})^2}}$$

$$\hat{r}_{ui} = \bar{r}_i + \frac{\sum_{i' \in K_i} sim(i, i') \cdot (r_{ui'} - \bar{r}_{i'})}{\sum_{i' \in K_i} |sim(i, i')|}$$

Kỹ thuật láng giềng

Giải pháp 1: User-base

```

1: procedure USERKNN-CF ( $\bar{r}_u, r, D^{train}$ )
2: for  $u=1$  to  $N$  do
3:   Tính  $Sim_{uu'}$ 
4: end for
5: Sort  $Sim_{uu'}$ 
6: for  $k=1$  to  $K$  do
7:    $K_u \leftarrow k$ 
8: end for
9: for  $i = 1$  to  $M$  do
10:  Tính  $\hat{r}_{ui}$ 
11: end for
12: end procedure
  
```

Giải pháp 2: Item-base

```

1. procedure ITEMKNN-CF ( $\bar{r}_u, r, D^{train}$ )
2. for  $i=1$  to  $M$  do
3.   Tính  $Sim_{ij}$ 
4. end for
5. Sort  $Sim_{ij}$ 
6. for  $i=1$  to  $K$  do
7.    $K_i \leftarrow k$ 
8. end for
9. for  $u = 1$  to  $N$  do
10.  Tính  $\hat{r}_{u,i}$ 
11. end for
12. end procedure
  
```

Vấn đề: COLD START PROBLEM

Kỹ thuật láng giềng

Vấn đề: COLD START PROBLEM

Giải pháp

$$\widehat{r}_{ui} = \alpha \widehat{r}_{ui}^i + (1 - \alpha) \widehat{r}_{ui}^u$$

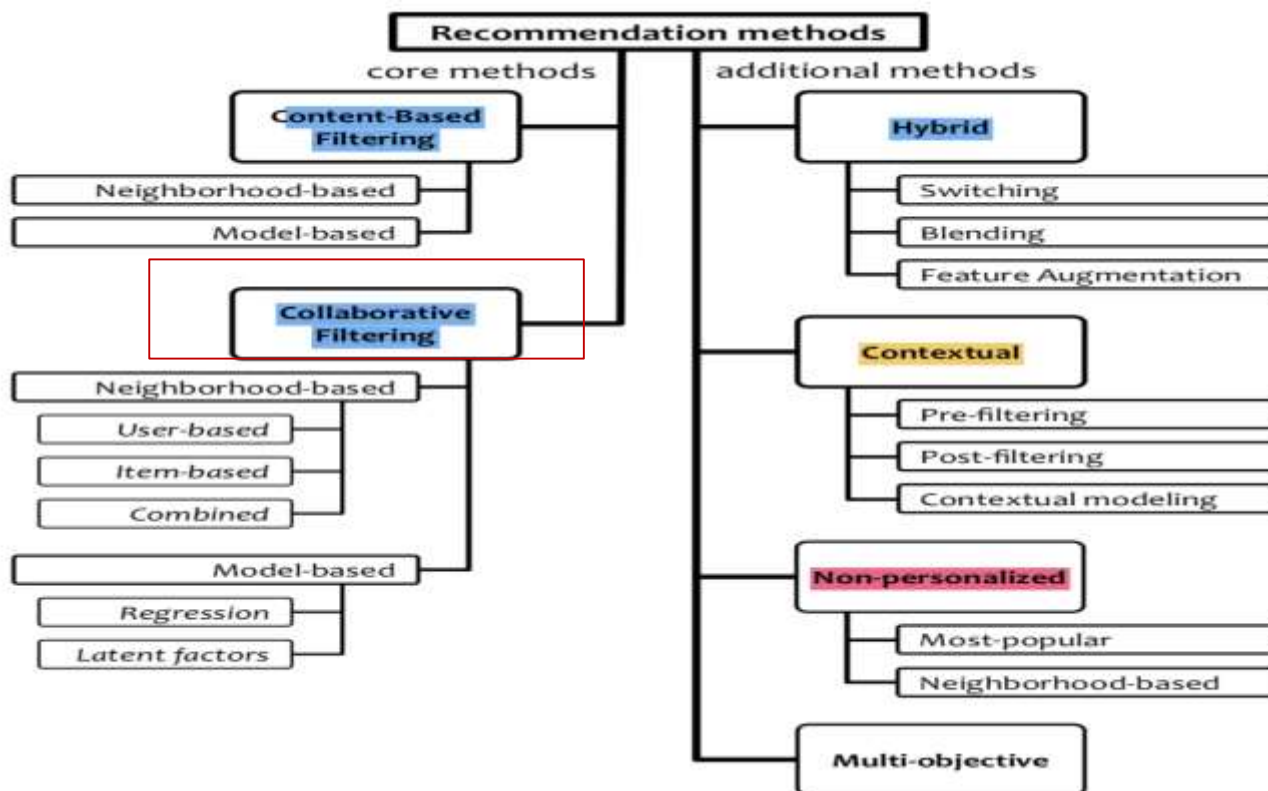
$$\hat{r}_{ui} = \mu + b_u + b_i$$

$$\mu = \frac{\sum_{(u,i,r) \in \mathcal{D}^{train}} r}{|\mathcal{D}^{train}|}$$

$$b_u = \frac{\sum_{(u',i,r) \in \mathcal{D}^{train} | u'=u} (r - \mu)}{|\{(u',i,r) \in \mathcal{D}^{train} | u'=u\}|}$$

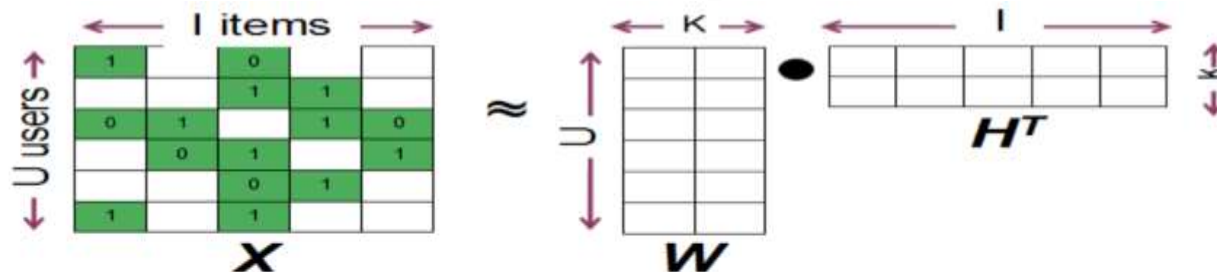
$$b_i = \frac{\sum_{(u,i',r) \in \mathcal{D}^{train} | i'=i} (r - \mu)}{|\{(u,i',r) \in \mathcal{D}^{train}\} | i'=i|}$$

Nhóm kỹ thuật chính



Mô hình nhân tố tiềm ẩn

- kỹ thuật phân rã (matrix factorization) là một dạng của mô hình nhân tố tiềm ẩn



Phân rã X thành 2 ma trận nhỏ hơn W và H sao cho ta có thể xây dựng lại X từ 2 ma trận con này

$$X \approx WH^T$$

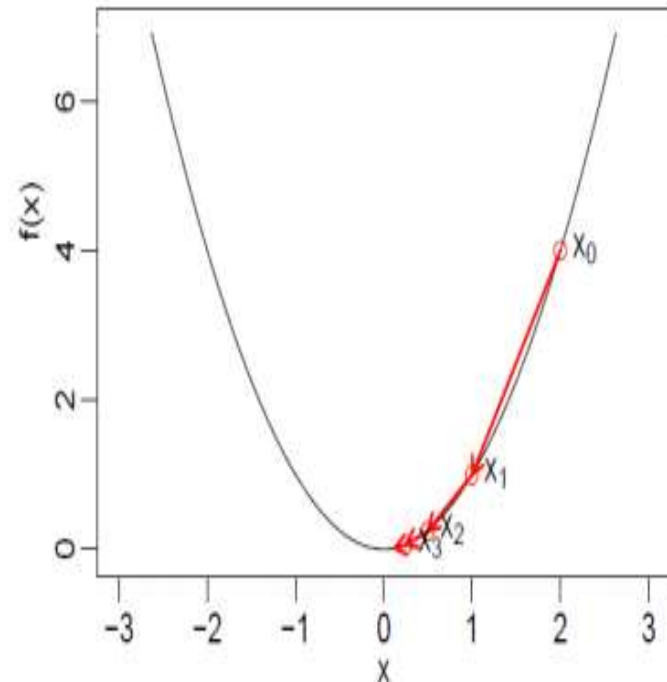
$$W \in \mathbb{R}^{|U| \times K}; \quad H \in \mathbb{R}^{|I| \times K};$$

K : là số nhân tố tiềm ẩn (latent factors) $K \ll |U|, K \ll |I|$

$$\mathcal{O}^{MF} = \sum_{(u,i) \in \mathcal{D}^{train}} (r_{ui} - \hat{r}_{ui})^2 = \sum_{(u,i) \in \mathcal{D}^{train}} \left(r_{ui} - \sum_{k=1}^K w_{uk} h_{ik} \right)^2$$

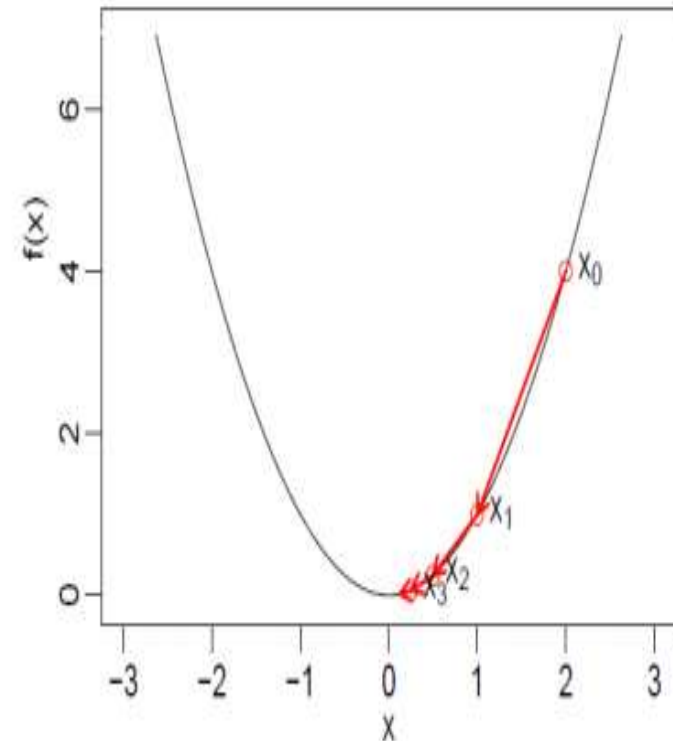
Kỹ thuật phân rã ma trận (matrix factorization)

- Tối ưu hóa bằng Gradient Descent
- Cho $f: \mathbb{R}^n \rightarrow \mathbb{R}$, tìm x sao cho $f(x)$ nhỏ nhất
- Ý tưởng:
 - Từ giá trị x_0
 - Cập nhật x_1 ,
 -
 - Cập nhật $x_{n+1} : f(x_{n+1}) \leq f(x_n)$



Kỹ thuật phân rã ma trận (matrix factorization)

- Tối ưu hóa bằng Gradient Descent
- Chọn hướng để cập nhật: $-\frac{\partial f}{\partial x}(x_n)$
- $x_{n+1} = x_n - \beta \cdot \frac{\partial f}{\partial x}(x_n)$



Kỹ thuật phân rã ma trận

$$O^{MF} = \sum_{(u,i) \in \mathcal{D}^{train}} (r_{ui} - \hat{r}_{ui})^2 = \sum_{(u,i) \in \mathcal{D}^{train}} \left(r_{ui} - \sum_{k=1}^K w_{uk} h_{ik} \right)^2$$

$$\frac{\partial}{\partial w_{uk}} O^{MF} = -2(r_{ui} - \hat{r}_{ui}) h_{ik}$$

$$\frac{\partial}{\partial h_{ik}} O^{MF} = -2(r_{ui} - \hat{r}_{ui}) w_{uk}$$

$$w_{uk}^{new} = w_{uk}^{old} - \beta \cdot \frac{\partial}{\partial w_{uk}} O^{MF} = w_{uk}^{old} + 2\beta \cdot (r_{ui} - \hat{r}_{ui}) h_{ik}$$

$$h_{ik}^{new} = h_{ik}^{old} - \beta \cdot \frac{\partial}{\partial h_{ik}} O^{MF} = h_{ik}^{old} + 2\beta \cdot (r_{ui} - \hat{r}_{ui}) w_{uk}$$

Kỹ thuật phân rã ma trận

$$O^{MF} = \sum_{(u,i) \in \mathcal{D}^{train}} (r_{ui} - \hat{r}_{ui})^2 = \sum_{(u,i) \in \mathcal{D}^{train}} \left(r_{ui} - \sum_{k=1}^K w_{uk} h_{ik} \right)^2$$

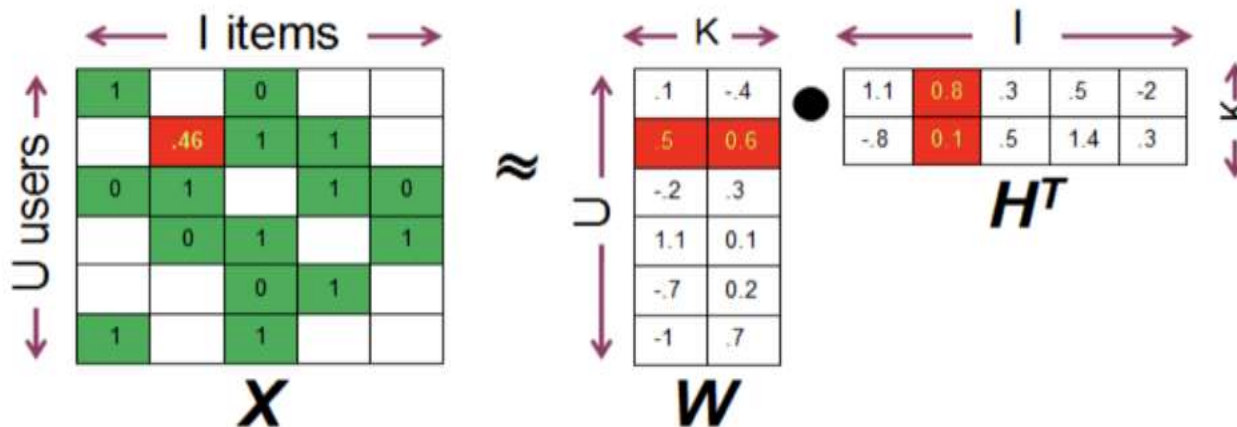
$$\frac{\partial}{\partial w_{uk}} O^{MF} = -2(r_{ui} - \hat{r}_{ui}) h_{ik}$$

$$\frac{\partial}{\partial h_{ik}} O^{MF} = -2(r_{ui} - \hat{r}_{ui}) w_{uk}$$

$$w_{uk}^{new} = w_{uk}^{old} - \beta \cdot \frac{\partial}{\partial w_{uk}} O^{MF} = w_{uk}^{old} + 2\beta \cdot (r_{ui} - \hat{r}_{ui}) h_{ik}$$

$$h_{ik}^{new} = h_{ik}^{old} - \beta \cdot \frac{\partial}{\partial h_{ik}} O^{MF} = h_{ik}^{old} + 2\beta \cdot (r_{ui} - \hat{r}_{ui}) w_{uk}$$

Kỹ thuật phân rã ma trận



$$\hat{r}_{ui} = \mathbf{w} \cdot \mathbf{h}^T = \sum_{k=1}^K w_{uk} h_{ik}$$

Kỹ thuật phân rã ma trận

- Vấn đề: "Quá khớp" (overfitting)
- Giải pháp: thêm vào hàm mục tiêu một đại lượng gọi là Regularization
- Hàm mục tiêu cần tối ưu bây giờ là:

$$\mathcal{O}^{MF} = \sum_{(u,i) \in \mathcal{D}^{train}} \left(r_{ui} - \sum_{k=1}^K w_{uk} h_{ik} \right)^2 + \lambda \cdot (||\mathbf{W}||_F^2 + ||\mathbf{H}||_F^2)$$

$\lambda \in (0..1)$ và $|| \cdot ||_F$ là chuẩn Frobenius:

$$||\mathbf{W}||_F = \sqrt{\sum_{u=1}^{|U|} \sum_{k=1}^K |w_{uk}|^2}$$

Kỹ thuật phân rã ma trận

- Vấn đề: "Quá khớp" (overfitting)
- Giải pháp: thêm vào hàm mục tiêu một đại lượng gọi là Regularization

Với hàm mục tiêu mới, cập nhật giá trị cho w_{uk} và h_{ik} bằng:

$$w_{uk}^{new} = w_{uk}^{old} - \beta \cdot \frac{\partial}{\partial w_{uk}} \mathcal{O}^{MF} = w_{uk}^{old} + \beta \cdot (2(r_{ui} - \hat{r}_{ui})h_{ik} - \lambda \cdot w_{uk})$$

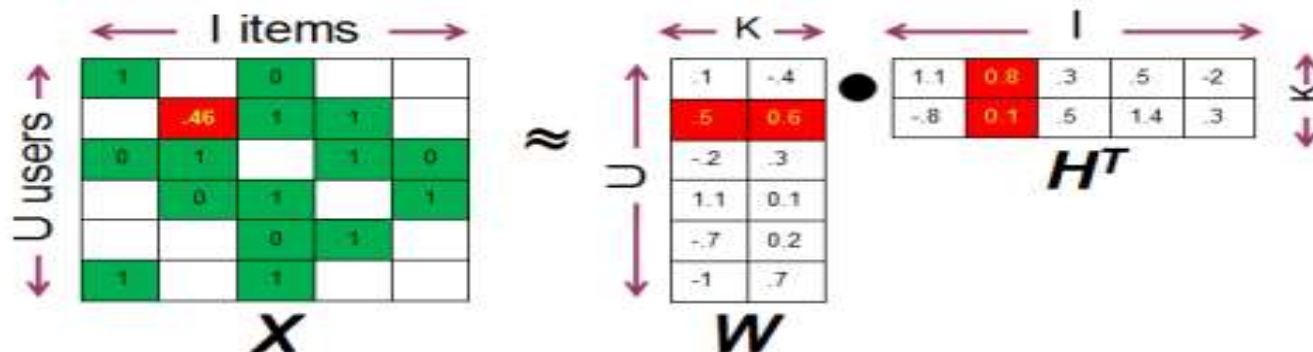
$$h_{ik}^{new} = h_{ik}^{old} - \beta \cdot \frac{\partial}{\partial h_{ik}} \mathcal{O}^{MF} = h_{ik}^{old} + \beta \cdot (2(r_{ui} - \hat{r}_{ui})w_{uk} - \lambda \cdot h_{ik})$$

Kỹ thuật phân rã ma trận

```

1: procedure MATRIXFACTORIZATION( $\mathcal{D}^{train}$ ,  $K$ ,  $\beta$ ,  $\lambda$ , stopping condition)
  // Let  $W[|U|][K]$  and  $H[|I|][K]$  be latent factors of users and items
2:    $W \leftarrow \mathcal{N}(0, \sigma^2)$ 
3:    $H \leftarrow \mathcal{N}(0, \sigma^2)$ 
4:   while (Stopping criterion is NOT met) do
5:     Draw randomly  $(u, i, r)$  from  $\mathcal{D}^{train}$ 
6:      $\hat{r} \leftarrow 0$ 
7:     for  $k \leftarrow 1, \dots, K$  do
8:        $\hat{r} \leftarrow \hat{r} + W[u][k] \cdot H[i][k]$ 
9:     end for
10:     $e_{ui} = r - \hat{r}$ 
11:    for  $k \leftarrow 1, \dots, K$  do
12:       $W[u][k] \leftarrow W[u][k] + \beta \cdot (e_{ui} \cdot H[i][k] - \lambda \cdot W[u][k])$ 
13:       $H[i][k] \leftarrow H[i][k] + \beta \cdot (e_{ui} \cdot W[u][k] - \lambda \cdot H[i][k])$ 
14:    end for
15:  end while
16:  return  $\{W, H\}$ 
17: end procedure
  
```

Cải tiến kỹ thuật phân rã ma trận



Hàm dự đoán

$$\hat{r}_{ui} = \mu + b_u + b_i + \sum_{k=1}^K w_{uk} h_{ik}$$

Cải tiến kỹ thuật phân rã ma trận

1. Procedure: ResultPrediction_BMF(D^{train} , K , β , λ , stopping condition)

Let $s \in S$ be a student, $i \in I$ a item, $p \in P$ a score

Let $W[S][K]$ and $H[I][K]$ be latent factors of students and tasks

Let $b_s[S]$ and $b_i[I]$ be students-bias and task-bias

$$2. \quad \mu \leftarrow \frac{\sum_{p \in D^{train}} p}{|D^{train}|}$$

3. for each student s do

$$4. \quad b_s[s] \leftarrow \frac{\sum_i (p_{si} - \mu)}{|D_s^{train}|}$$

5. end for

6. for each task i do

$$7. \quad b_i[i] \leftarrow \frac{\sum_u (p_{ui} - \mu)}{|D_i^{train}|}$$

8. end for

$$9. \quad W \leftarrow N(0, \sigma^2)$$

$$10. \quad H \leftarrow N(0, \sigma^2)$$

11. while (Stopping criterion is NOT met) do

12. Draw randomly (s, i, p_{si}) from D^{train}

$$13. \quad \hat{p}_{si} \leftarrow \mu + b_s[s] + b_i[i] + \sum_k (W[s][k] * H[i][k])$$

$$14. \quad e_{si} = p_{si} - \hat{p}_{si}$$

$$15. \quad \mu \leftarrow \mu + \beta * e_{si}$$

$$16. \quad b_s[s] \leftarrow b_s[s] + \beta * (e_{si} - \lambda * b_s[s])$$

$$17. \quad b_i[i] \leftarrow b_i[i] + \beta * (e_{si} - \lambda * b_i[i])$$

18. for $k \leftarrow 1, \dots, K$ do

$$19. \quad W[s][k] \leftarrow W[s][k] + \beta * (2e_{si} * H[i][k] - \lambda * W[s][k])$$

$$20. \quad H[i][k] \leftarrow H[i][k] + \beta * (2e_{si} * W[s][k] - \lambda * H[i][k])$$

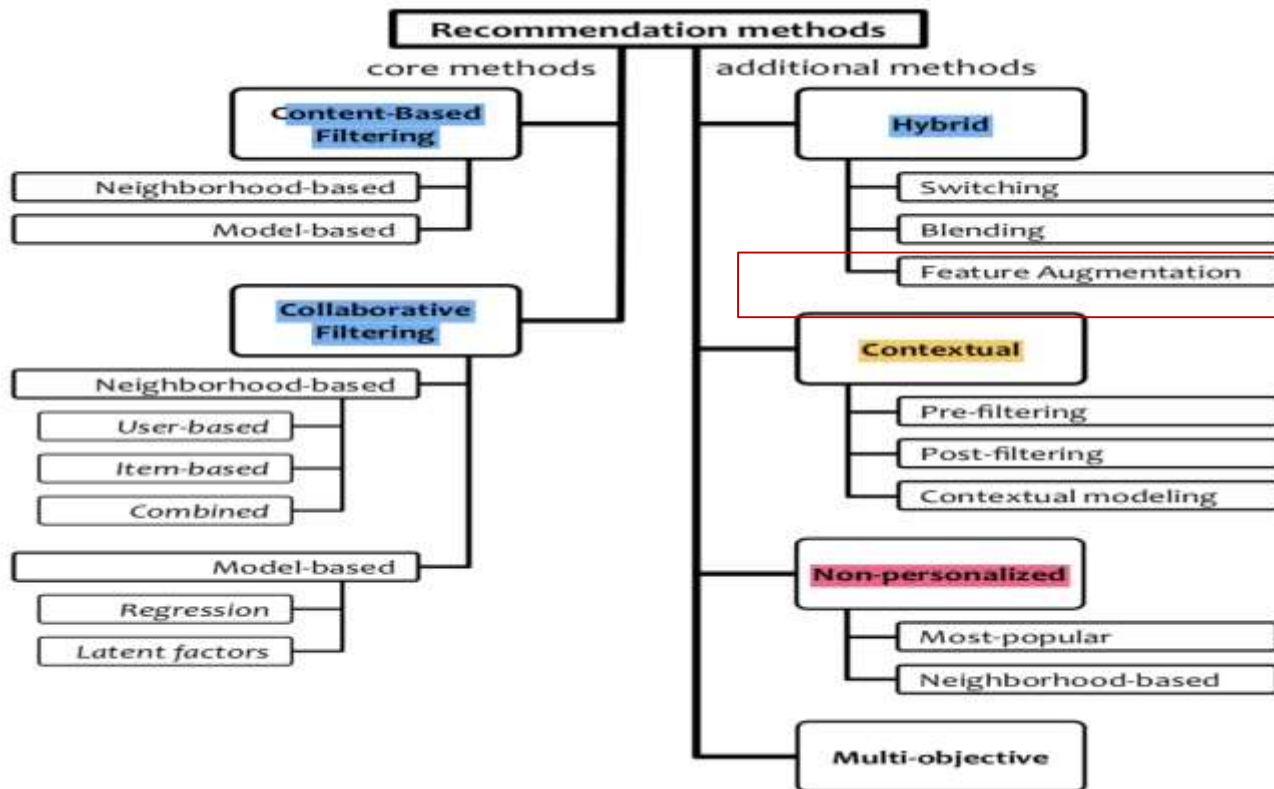
21. end for

22. end while

23. return $\{W, H, b_s, b_i, \mu\}$

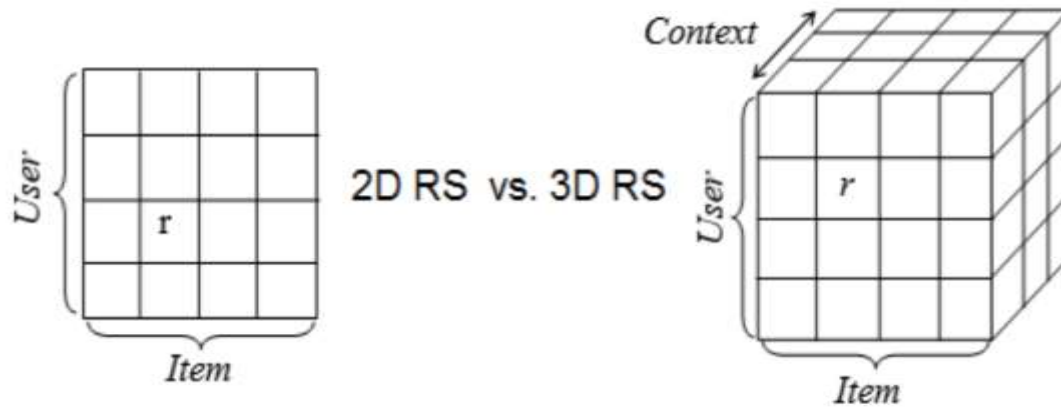
24. end procedure.

Nhóm kỹ thuật RS chính



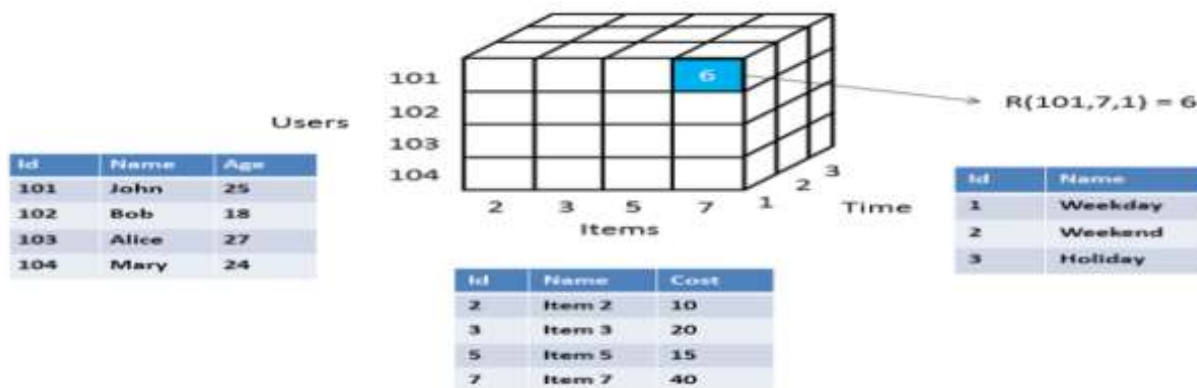
Gợi ý theo ngữ cảnh

- 2D RS: $U \times I \rightarrow R$
- 3D RS: $U \times I \times C \rightarrow R$



Gợi ý theo ngữ cảnh

- Dữ liệu ngữ cảnh (contextual data)



- Ngữ cảnh có thể:
 - Mùa
 - Thời gian
 - Nhiệt độ
 - Vị trí
 - Bạn đồng hành
- Ngữ cảnh của aki-travel?

[Adomavicius and Tuzhilin, 2011]

Gợi ý theo ngữ cảnh

- RS: Pre-filtering

user	item	time	Bạn Đồng hành	Thời tiết	rate
1	2	Cuối tuần	Bạn bè	Trời nắng	4
1	5	Cuối tuần	Một mình	Trời âm u	1
1	3	Lễ - tết	Gia đình	Trời trong xanh	5
2	2	Ngày trong tuần	Bạn bè	Trời nắng	2
2	1	Lễ - tết	Gia đình	Trời trong xanh	3
3	5	Lễ - tết	Gia đình	Trời trong xanh	4
3	4	Cuối tuần	Bạn bè	Trời nắng	3
4	3	Lễ - tết	Gia đình	Trời trong xanh	5

user	item	rate
1	3	5
2	1	3
3	5	4
4	3	5

Gợi ý theo ngữ cảnh

- RS: Pre-filtering
- 2D:

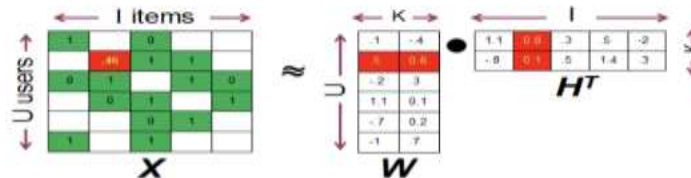
Biased Matrix Factorization:

$$\hat{r}_{ui} = \mu + b_u + b_i + w_u \cdot h_i^T$$

SVD++ [Koren, 2008]:

$$\hat{r}_{ui} = \mu + b_u + b_i + \left(w_u + \frac{1}{\sqrt{|N(u)|}} \sum_{j \in N(u)} y_j \right) \cdot h_i^T$$

where $N(u)$ is the set of implicit information by user u



$R < 20 \text{ km}$, thì $B \gg E \gg F \gg G, \dots$

Gợi ý theo ngữ cảnh

```

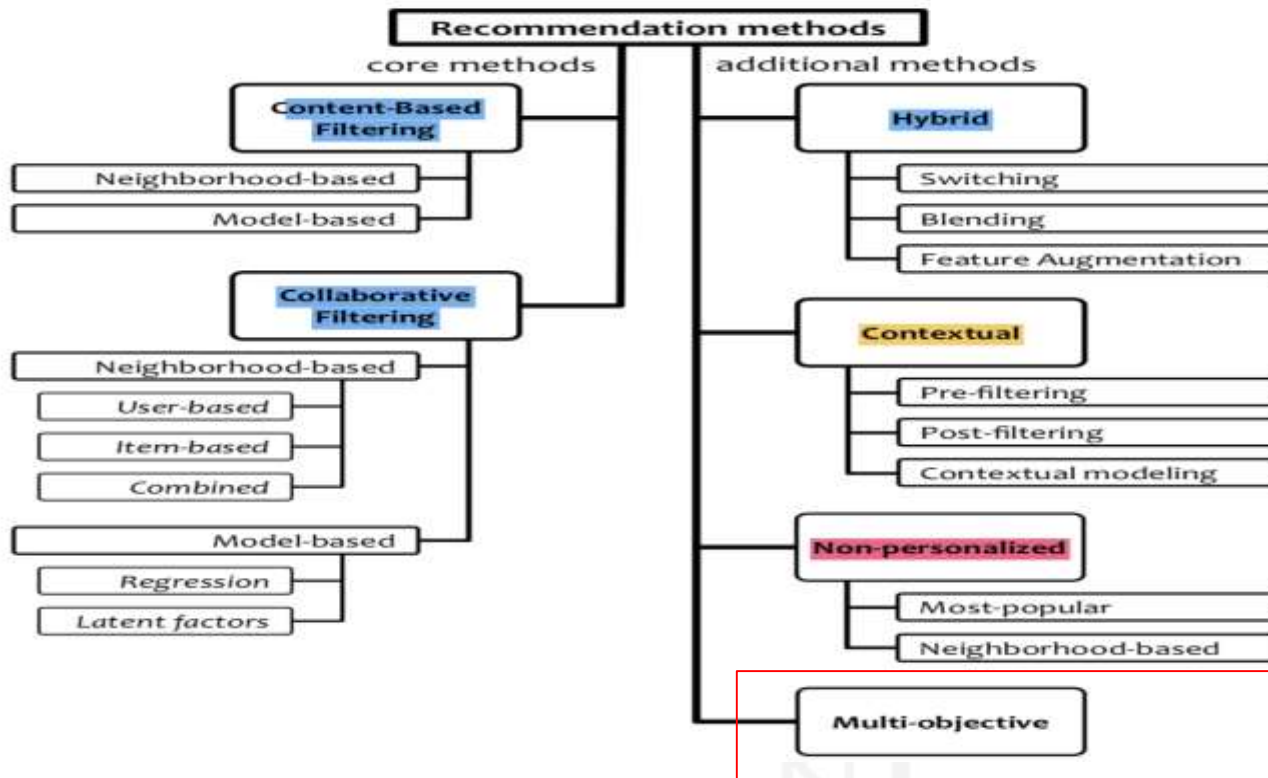
1: procedure ContextAware-MF ( $D^{\text{Train}}$ , Iter, K,  $\beta$ ,  $\lambda$ )
//  $W[[U]][[K]]$  và  $H[[I]][[K]]$  là 2 tham số cần tìm
2:  $W := N(0, \sigma^2)$  //khởi tạo giá trị theo phân phối chuẩn
3:  $H := N(0, \sigma^2)$  //khởi tạo giá trị theo phân phối chuẩn
4:  $D^{\text{TrainC}} = \text{Pre-filtering}(D^{\text{Train}})$ 
5: for (iter:=1; iter <= Iter *  $|D^{\text{TrainC}}|$ ; iter++)
6:   Chọn ngẫu nhiên một dòng (u, i,  $r_{ui}$ ) từ  $D^{\text{TrainC}}$ 
7:    $\hat{r}_{ui} := 0$ 
8:   for (k:=1; k<=K; k++)
9:      $\hat{r}_{ui} := \hat{r}_{ui} + W[u][k] * H[i][k]$ 
10:   end for
11:    $e_{ui} = r_{ui} - \hat{r}_{ui}$ 
12:   for (k:=1; k<=K; k++)
13:      $W[u][k] := W[u][k] + \beta * (e_{ui} * H[i][k] - \lambda * W[u][k])$ 
14:      $H[i][k] := H[i][k] + \beta * (e_{ui} * W[u][k] - \lambda * H[i][k])$ 
15:   end for
16:   Break nếu đã hội tụ
17: end for
18: return {W, H}
19: Post-filtering(Tập kết quả được dự đoán dùng W, H)
20: end procedure
  
```

- Vấn đề:

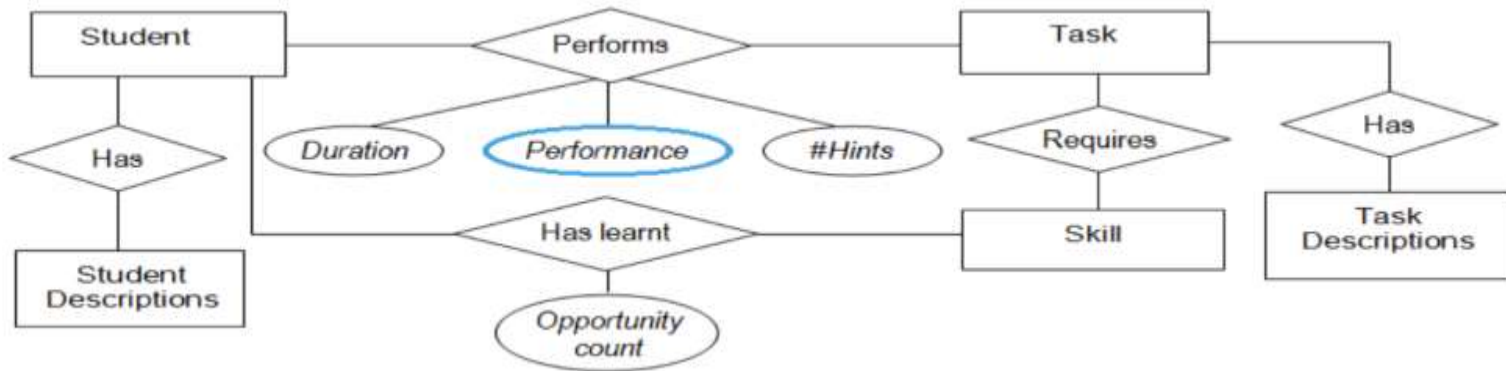
- + Di chuyển với tốc độ cao
- + Tốc độ mạng chậm

- Giải pháp ?

Nhóm kỹ thuật chính



Multi_RM_F



← Tasks →

↑ Students ↓

p		p		
		p	p	
p	p		p	p
	p	p		p
		p	p	
p		p		

a. Student-Performs-Task

← Skills →

↑ Tasks ↓

1			1	
	1			
			1	
1				
		1	1	
	1			1

b. Task-Requires-Skill

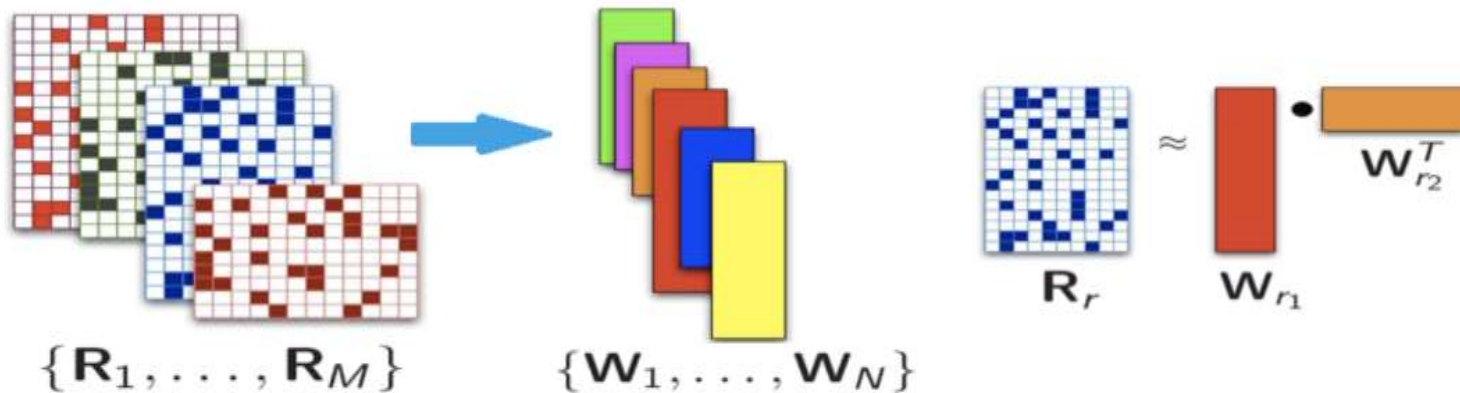
← Skills →

↑ Students ↓

1			2	
1			1	
2	1	1	2	
	1	1	2	
1			1	
1			2	

c. Student-hasLearnt-Skill

Multi_RMF



Picture: adapted from [Lippert et al., 2008]

Objective function:

$$\mathcal{O}^{\text{MRMF}} = \sum_{r=1}^M \sum_{(u,i) \in R_r} ((R_r)_{ui} - \mathbf{w}_{r1u} \cdot \mathbf{w}_{r2i}^T)^2 + \lambda \left(\sum_{j=1}^N \|\mathbf{w}_j\|_F^2 \right)$$

[Lippert et al., 2008, Singh and Gordon, 2008]

Multi_RMFB

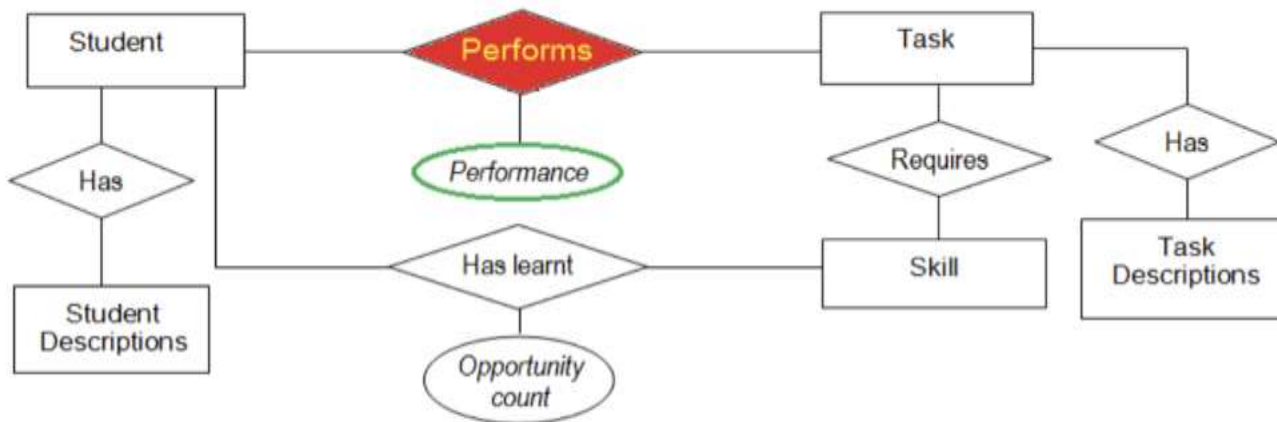
MF objective function:

$$\begin{aligned}\mathcal{O}^{MF} &= \sum_{(u,i) \in \mathcal{D}^{train}} \left(r_{ui} - \mathbf{w}_u \cdot \mathbf{h}_i^T \right)^2 + \lambda (\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2) \\ &= \sum_{(u,i) \in \mathcal{D}^{train}} \left(r_{ui} - \sum_{k=1}^K w_{uk} h_{ik} \right)^2 + \lambda (\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2)\end{aligned}$$

MRMF objective function:

$$\mathcal{O}^{MRMF} = \sum_{r=1}^M \sum_{(u,i) \in \mathbf{R}_r} \left((\mathbf{R}_r)_{ui} - \mathbf{w}_{r_1 u} \cdot \mathbf{w}_{r_2 i}^T \right)^2 + \lambda \left(\sum_{j=1}^N \|\mathbf{w}_j\|_F^2 \right)$$

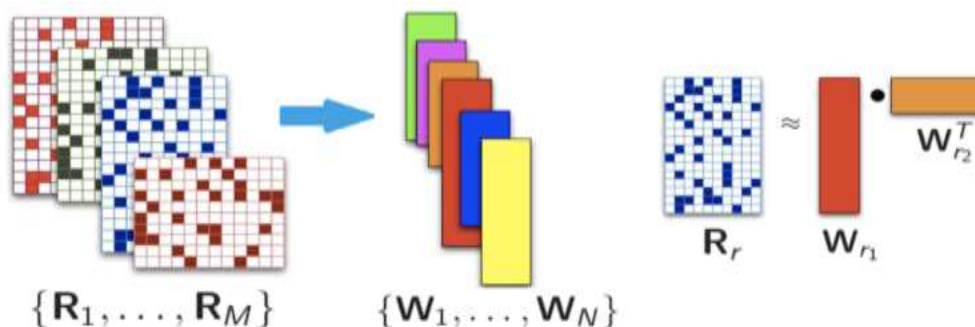
Multi- RMF có trọng số



Quan hệ chính (chứa thuộc tính cần dự đoán) có thể quan trọng hơn những quan hệ khác.

⇒ **Cần có trọng số cao hơn**

Multi_RMFM có trọng số



Hàm mục tiêu:

$$\mathcal{O}^{\text{WMMRF}} = \sum_{r=1}^M \Theta_r \sum_{(s,i) \in R_r} \left((R_r)_{si} - \mathbf{w}_{r1s} \cdot \mathbf{w}_{r2i}^T \right)^2 + \lambda \left(\sum_{j=1}^N \|\mathbf{w}_j\|_F^2 \right)$$

Hàm trọng số:

$$\Theta_r = \begin{cases} 1, & \text{nếu } r \text{ là quan hệ chính} \\ \theta, & \text{nếu ngược lại } (0 < \theta \leq 1) \end{cases}$$

Trong trường hợp xấu nhất ($\theta = 1$), WMMRF vẫn tương đương với MRMF

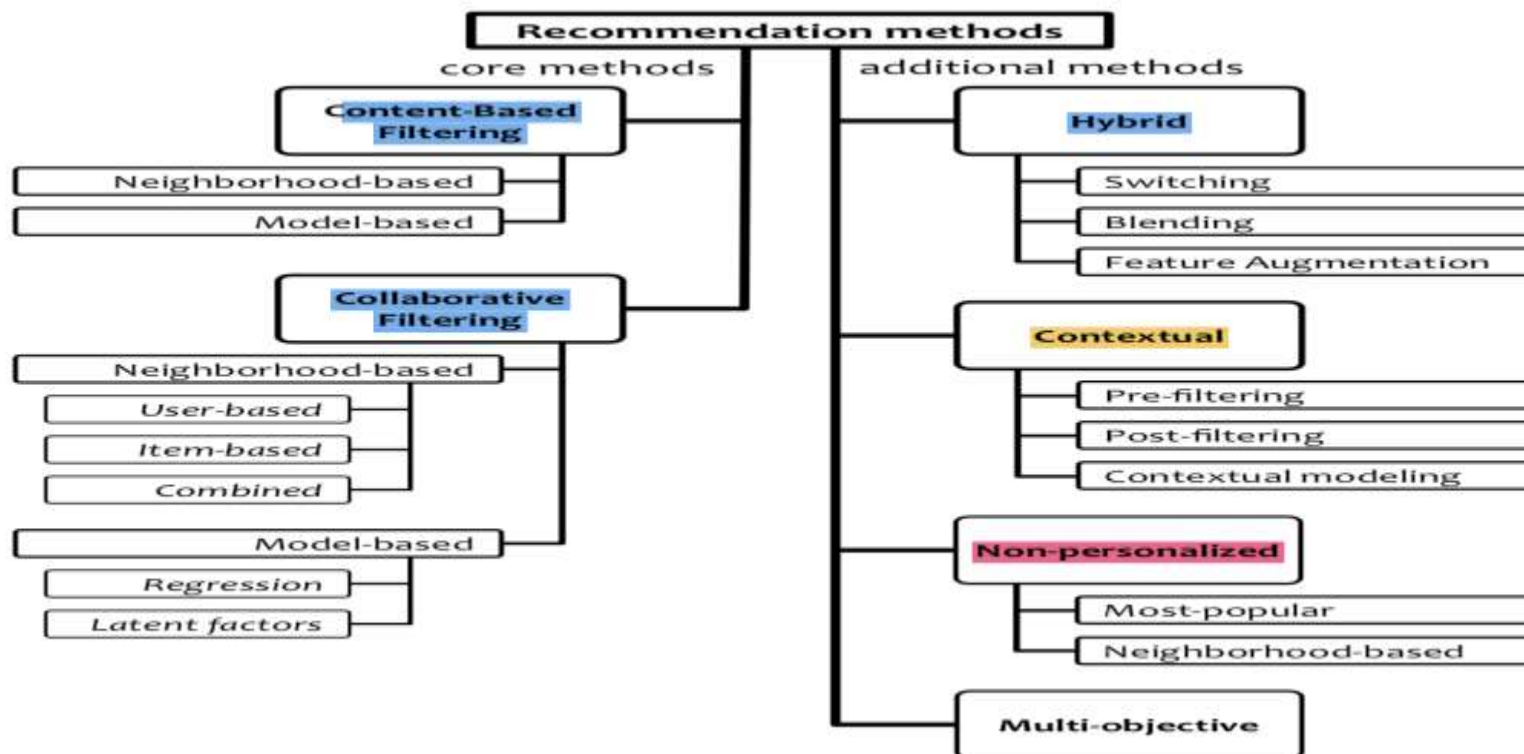
Multi_RMFM có trọng số

```

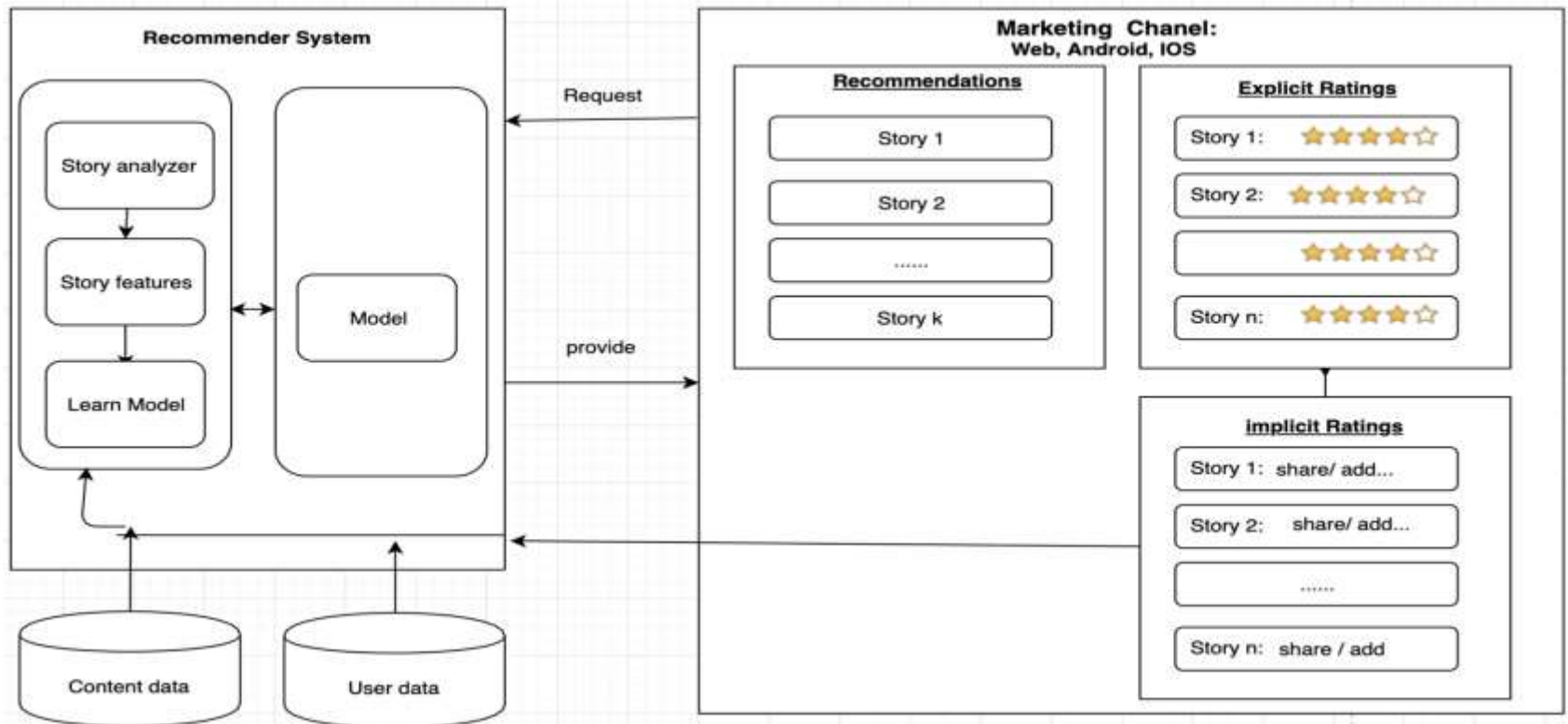
1: procedure LEARNWMMFM(
     $E_1, \dots, E_N$ : Thực thể;  $R_1, \dots, R_M$ : Quan hệ;  $\lambda$ :
    hằng số chính tắc hóa (regularization);  $\beta$ : Tốc độ học;
     $K$ : Số nhân tố tiềm ẩn;  $\theta$ : trọng số; Điều kiện dừng)
2:   for  $j \leftarrow 1 \dots N$  do
3:      $W_j \leftarrow$  Rút ngẫu nhiên từ  $\mathcal{N}(\mu, \sigma^2)$ 
4:   end for
5:   for  $r \leftarrow 1 \dots M$  do
6:     Khởi tạo  $\Theta_r$  sử dụng công thức (6)
7:   end for
8:   while (Điều kiện dừng chưa thỏa) do
9:     for mỗi quan hệ  $R_r = \{(E_{1r}; E_{2r})\}$  in
        $\{R_1, \dots, R_M\}$  do
10:      for  $l \leftarrow 1 \dots |R_r|$ , do
11:        Lấy ngẫu nhiên bộ  $(s, i)$  trong  $R_r$ 
12:         $w_{r1s} \leftarrow w_{r1s} - \beta \left( \frac{\partial \mathcal{O}^{WMMFM}}{\partial w_{r1s}} \right)$ 
13:         $w_{r2i} \leftarrow w_{r2i} - \beta \left( \frac{\partial \mathcal{O}^{WMMFM}}{\partial w_{r2i}} \right)$ 
14:      end for
15:    end for
16:  end while
17:  return  $\{W_j\}_{j=1 \dots N}$ 
18: end procedure

```

Tóm tắt



Trao đổi



<https://github.com/IBM/elasticsearch-spark-recommender/blob/master/README.md#links>

Cảm Ơn!