02

collaborative approaches

and

Content based approaches

HUYINIT

https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada

Recommender systems

Content based methods

Define a model for user-item interactions where users and/or items representations are given (explicit features).

Collaborative filtering methods

Model based

Define a model for user-item interactions where users and items representations have to be learned from interactions matrix.

Memory based

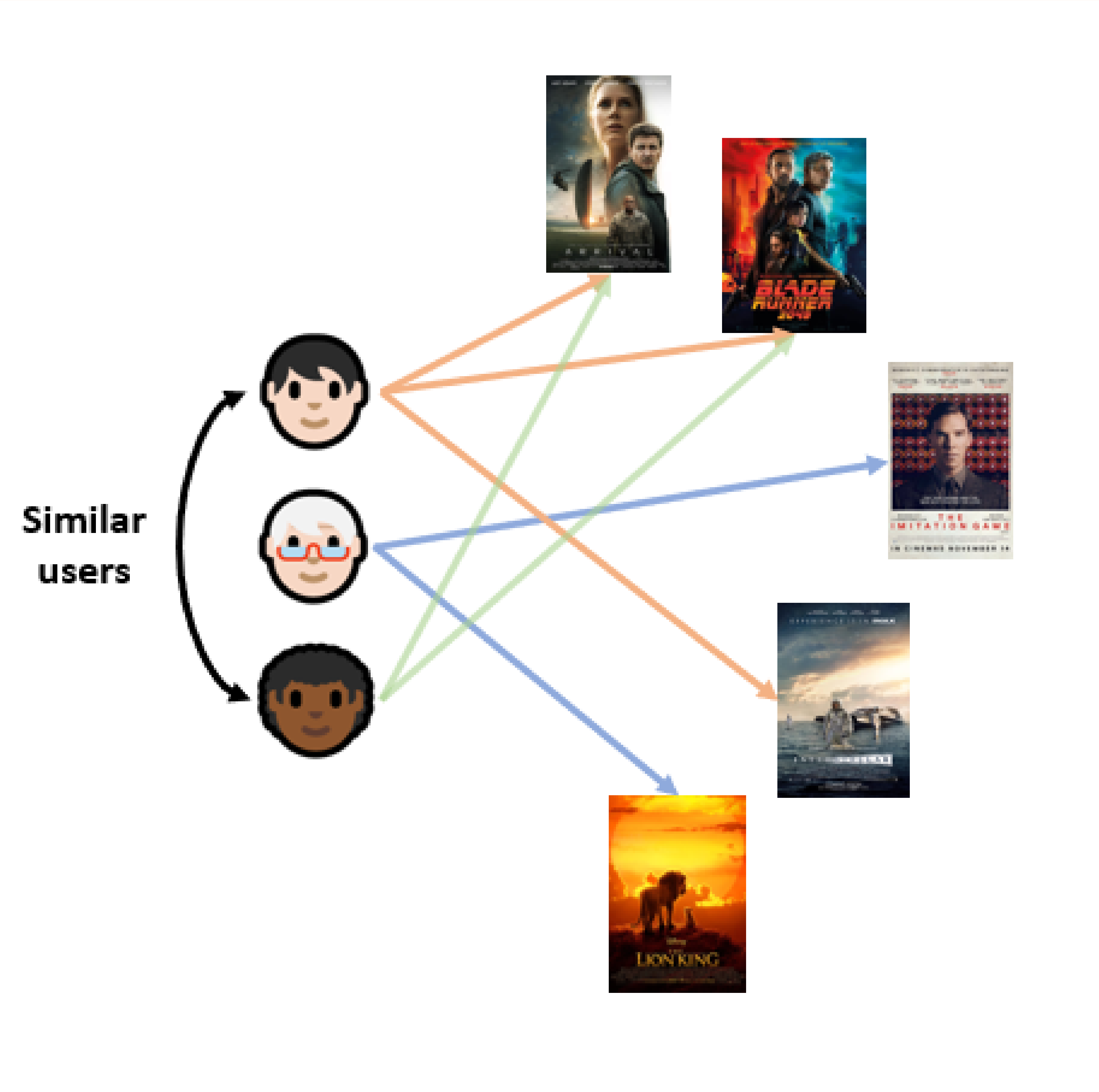
Define no model for user-item interactions and rely on similarities between users or items in terms of observed interactions.

Hybrid methods

Mix content based and collaborative filtering approaches.

Memory based collaborative approaches

User-user

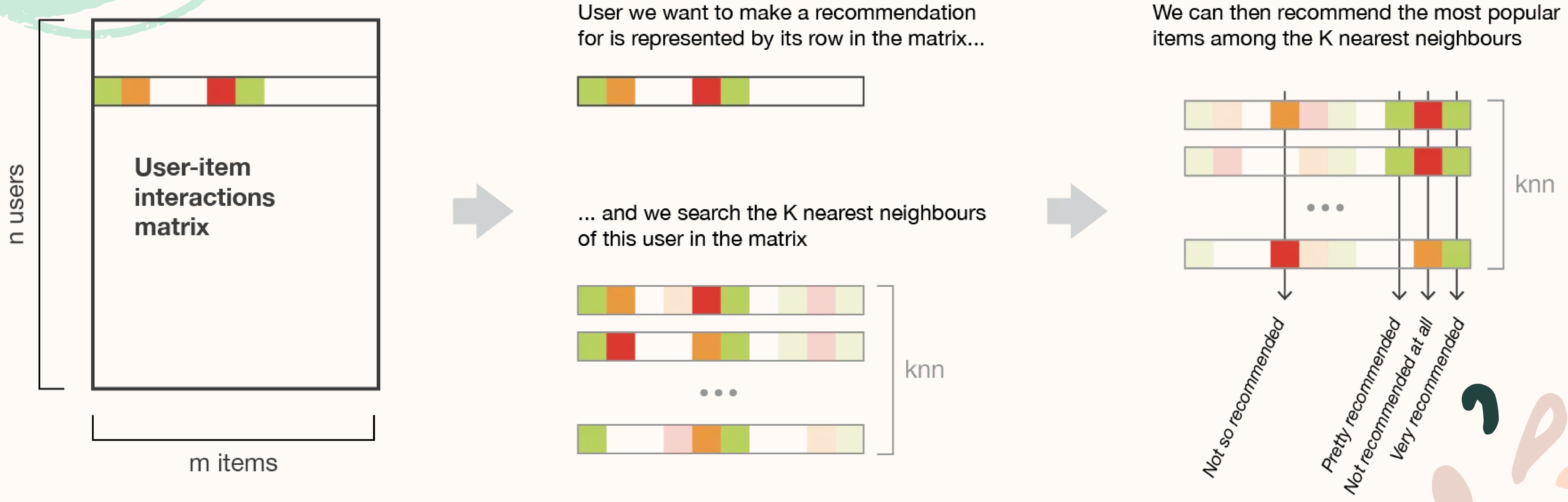


User-user

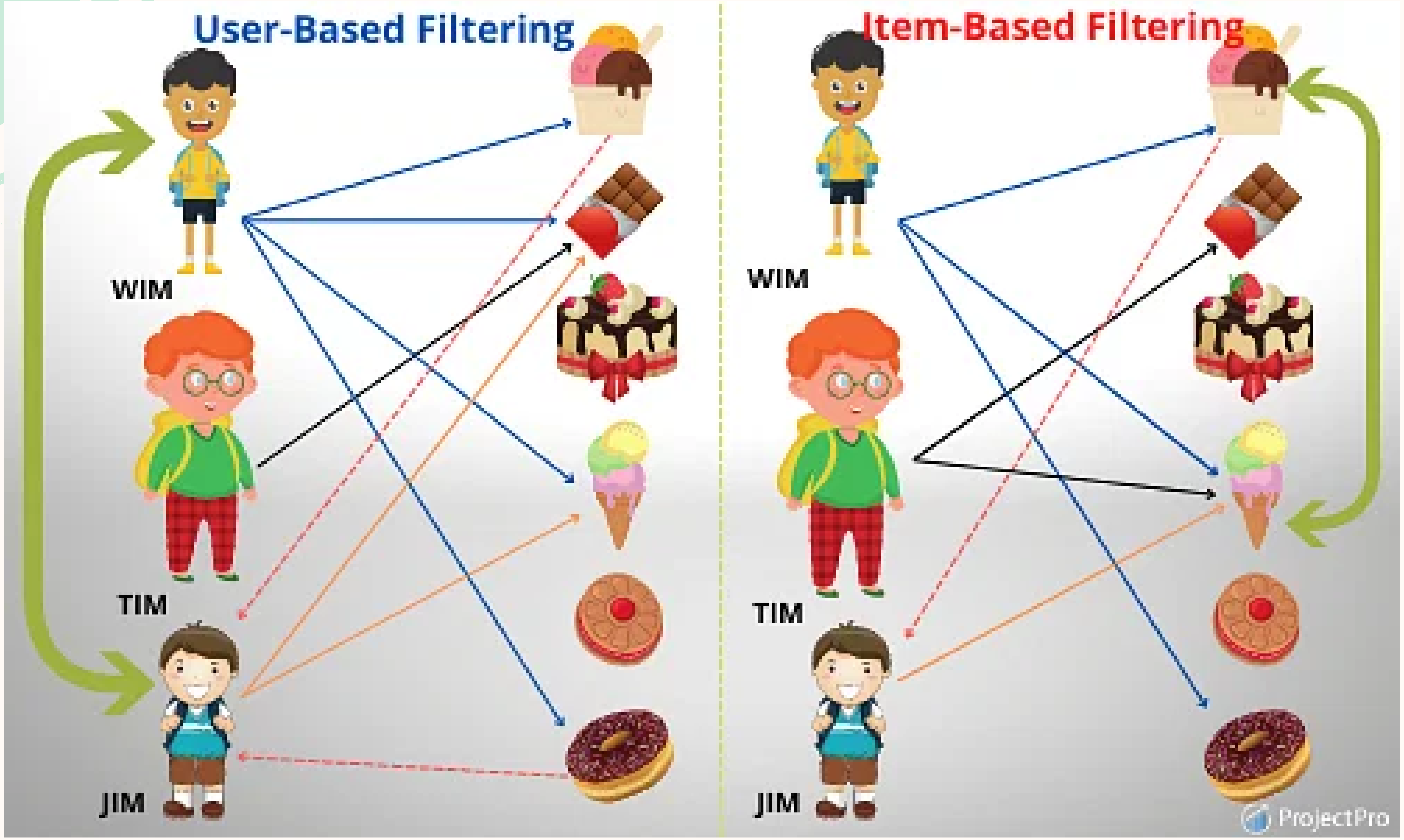
positive interactions

neutral interactions

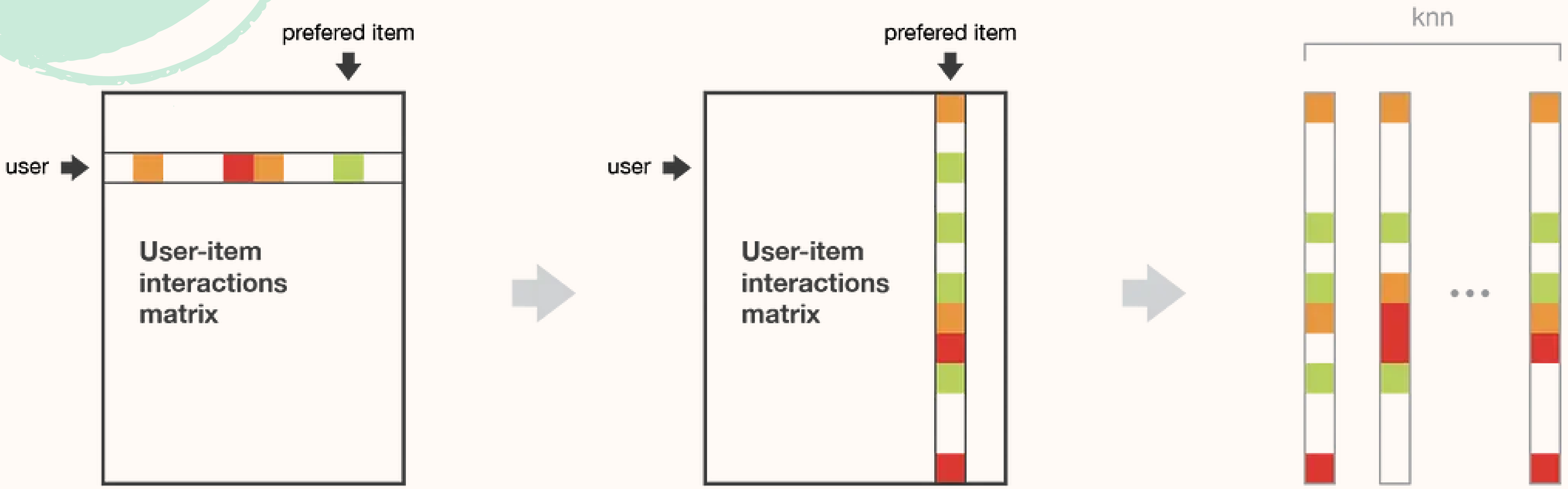
negative interactions



Item-item


User-Based Filtering
WIM
TIM
JIM
Item-Based Filtering
WIM
TIM
JIM
ProjectPro

Item-item

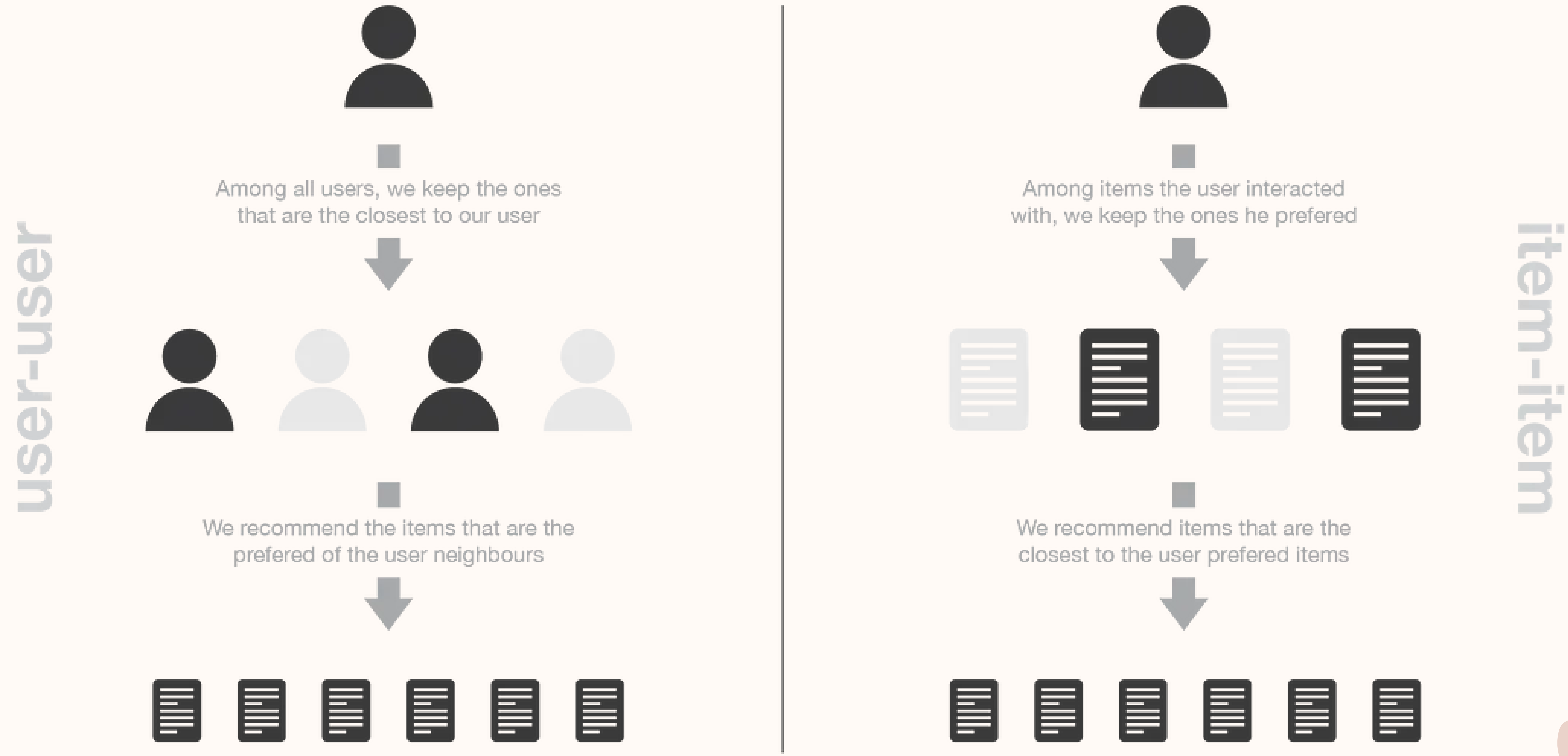


We identify the prefered item of user we want to make recommendation for.

The prefered item is represented by its column in the matrix.

We can search and recommend the K nearest items to this "prefered item"

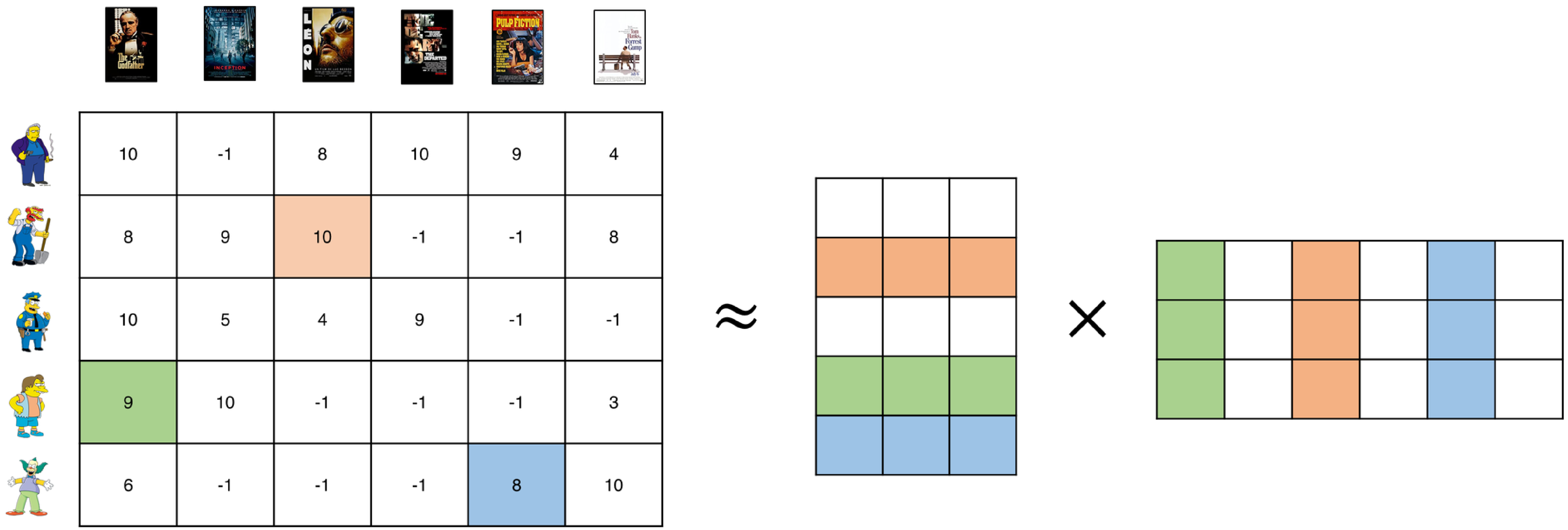
So sánh user-user và item-item



Độ phức tạp và tác dụng phụ

Collaborative filtering techniques	Memory-based filtering	Model-based filtering	Hybrid filtering
Advantages	Easy implementation	Better sparsity and scalability, Improve recommendati on performance, Not require so much memory and cpu time	Improve sparsity
Disadvantages	Performance decrease when data are sparse, require lots of memory and cpu time	Loose useful information for dimensionalit y reduction technique, Expensive	Increased complexity, expensive in implementatio n

Model based collaborative approaches



10	-1	8	10	9	4
8	9	10	-1	-1	8
10	5	4	9	-1	-1
9	10	-1	-1	-1	3
6	-1	-1	-1	8	10

User-item Interaction Matrix (R)

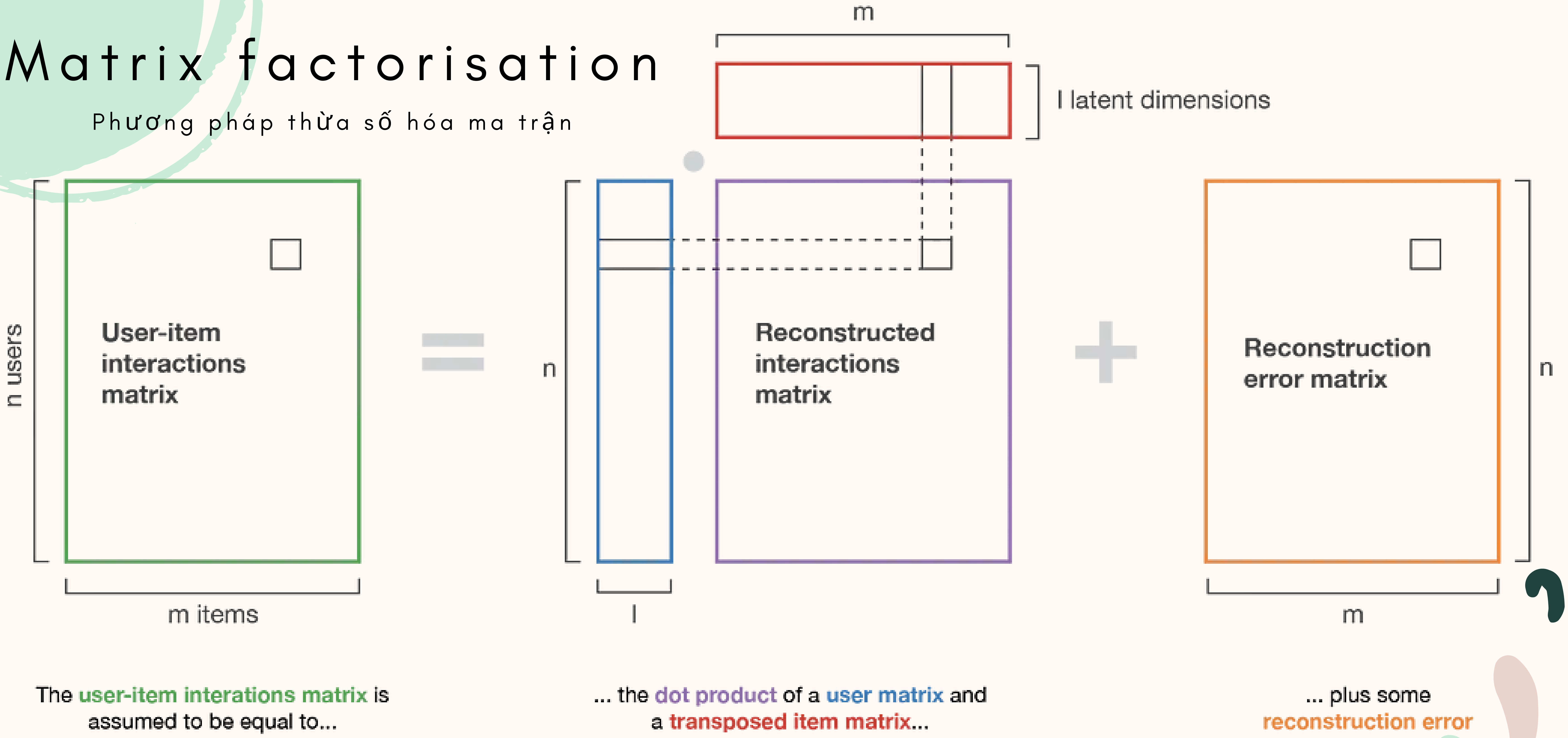
User Matrix (Q)

Item Matrix (P)

Model based collaborative approaches

Matrix factorisation

Phương pháp thừa số hóa ma trận



Model based collaborative approaches

Mathematics of matrix factorisation

$$M \approx X.Y^T$$

interaction matrix M (nxm)

X is the "user matrix" (nxl)

Y is the "item matrix" (mxl)

$$user_i \equiv X_i \qquad \forall i \in \{1, ..., n\}$$

$$item_j \equiv Y_j \qquad \forall j \in \{1, ..., m\}$$

Model based collaborative approaches

$$(X,Y) = \underset{X,Y}{argmin} \sum_{(i,j)\in E} [(X_i)(Y_j)^T - M_{ij}]^2$$

E the ensemble of pairs (i,j) such that M_ij is set (not None),

we want to find X and Y that minimise the
"rating reconstruction error"

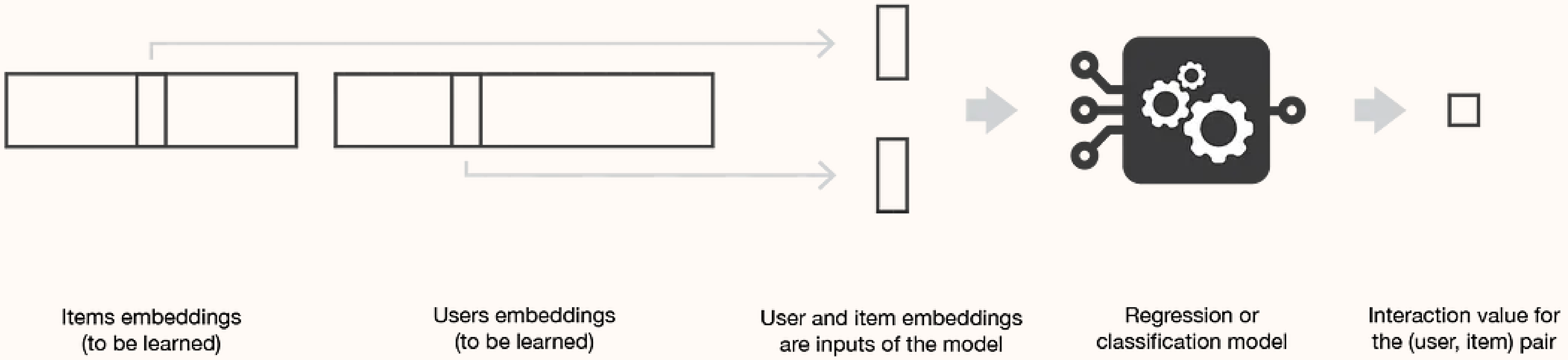
$$(X,Y) = \underset{X,Y}{argmin} \frac{1}{2} \sum_{(i,j)\in E} [(X_i)(Y_j)^T - M_{ij}]^2 + \frac{\lambda}{2}(\sum_{i,k}(X_{ik})^2 + \sum_{j,k}(Y_{jk})^2)$$

Model based collaborative approaches

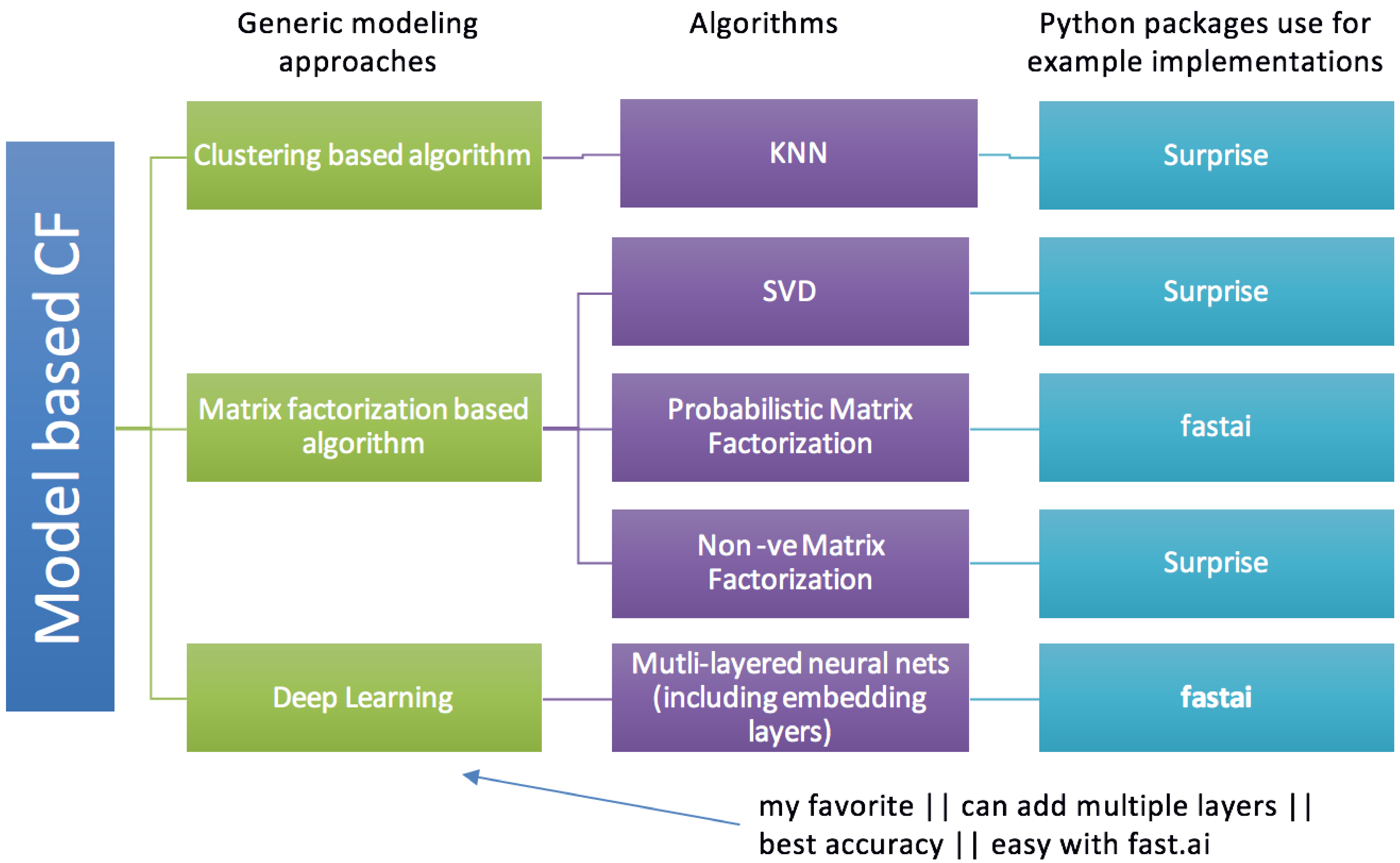
Extensions of matrix factorisation

$$(X,Y) = \underset{X,Y}{argmin} \frac{1}{2} \sum_{(i,j)\in E} [f((X_i)(Y_j)^T) - M_{ij}]^2 + \frac{\lambda}{2}(\sum_{i,k}(X_{ik})^2 + \sum_{j,k}(Y_{jk})^2)$$

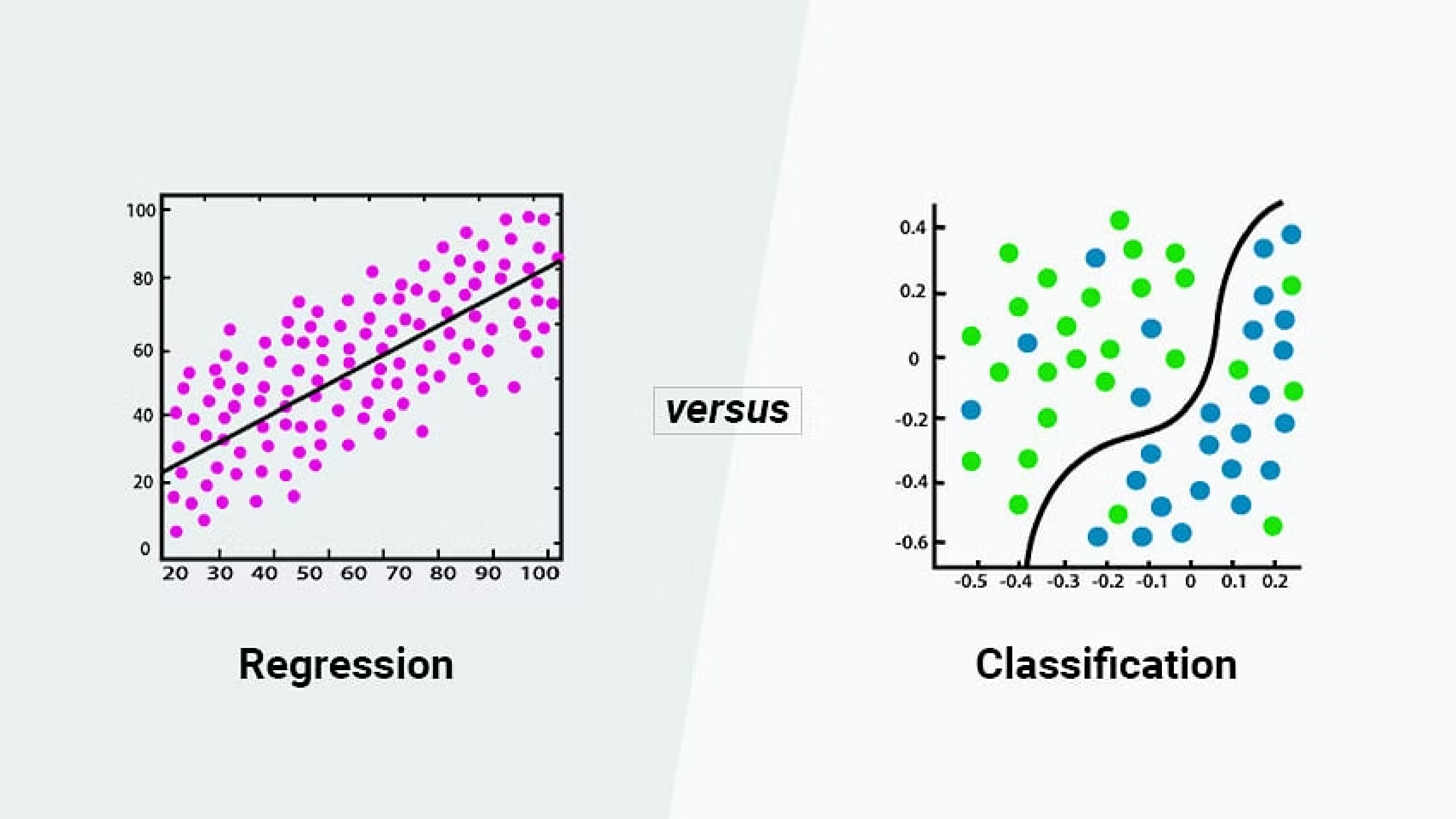
we get a model that takes its value in [0, 1]

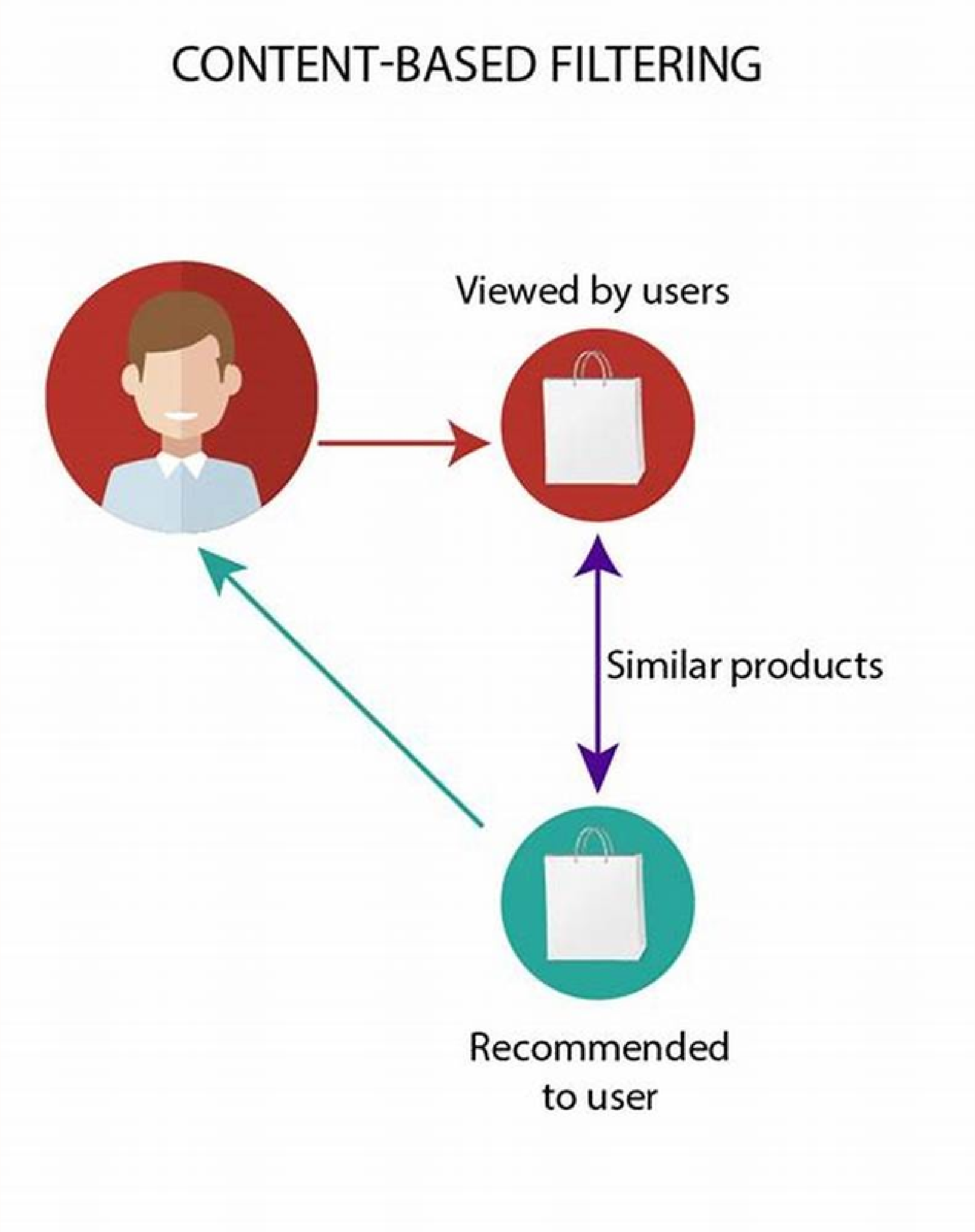


Model based collaborative approaches



Content based approaches





Content based approaches

Model for a given user based on items features

Items that the concerned user has interacted with (dataset)

item-centred

Model for a given item based on user features

Users that have interacted with the concerned item (dataset)

user-centred

Content based approaches
Item-centred Bayesian classifier

$$\frac{\mathbb{P}_{item}(like|user\_features)}{\mathbb{P}_{item}(dislike|user\_features)}$$

$$\mathbb{P}_{item}(like|user\_features) = \frac{\mathbb{P}_{item}(user\_features|like) \times \mathbb{P}_{item}(like)}{\mathbb{P}_{item}(user\_features)}$$

$$\mathbb{P}_{item}(dislike|user\_features) = \frac{\mathbb{P}_{item}(user\_features|dislike) \times \mathbb{P}_{item}(dislike)}{\mathbb{P}_{item}(user\_features)}$$

$$\frac{\mathbb{P}_{item}(like|user\_features)}{\mathbb{P}_{item}(dislike|user\_features)} = \frac{\mathbb{P}_{item}(user\_features|like) \times \mathbb{P}_{item}(like)}{\mathbb{P}_{item}(user\_features|dislike) \times \mathbb{P}_{item}(dislike)}$$

$\mathbb{P}_{item}(like)$ and $\mathbb{P}_{item}(dislike)(= 1 - \mathbb{P}_{item}(like))$

Content based approaches

User described by some features

(features can be of various kind
and define the inputs of the model)

Bayesian classifier for a given item

(parameters of the bayesian classifier
are specific to the item and learned
on past item interactions)

Predicted class ("like" or "dislike")

(output of the bayesian classifier model
when inputs are the features of the user)

Illustration of the item-centred content
based Bayesian classifier.

Content based approaches

User-centred linear regression

$$X_i = \underset{X_i}{argmin} \frac{1}{2} \sum_{(i,j) \in E} [(X_i)(Y_j)^T - M_{ij}]^2 + \frac{\lambda}{2} (\sum_k (X_{ik})^2)$$

for a given user i

Item described by some features

(features can be of various kind and define the inputs of the model)

Linear regression for a given user

(parameters of the linear regression are specific to the user and learned on past user interactions)

Predicted rating

(output of the linear regression model when inputs are the features of the item)

Evaluation of a recommender system

Performance Evaluation of Recommender Systems

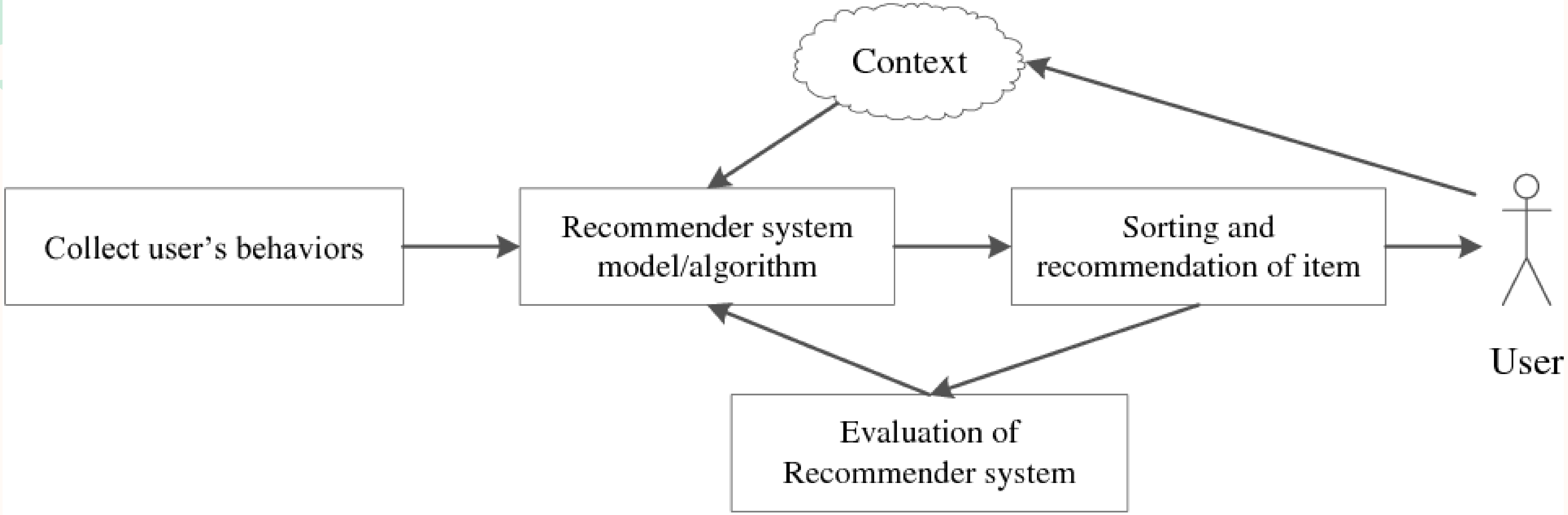


Figure 1. The module diagram for recommender system's design and evaluation

Metrics based evaluation

Human based evaluation

Thank you