Sistemas de Recomendação

TensorFlow

Motivações e Aplicações

- “Os SR provaram ser uma ótima abordagem para lidar com o problema de sobrecarga de informações citado” (RICCI; ROKACH; SHAPIRA, 2011)
- “1/3 dos consumidores que notam as recomendações acabam comprando algo baseado nelas” (GROSSMAN, 2013)
- SRs são sistemas complexos
- Informações sobre itens x usuários

Motivações e Aplicações

- Motivos para implantar um SR
  - Aumentar o número de itens vendidos
  - Vender itens mais diversificados
  - Aumentar a satisfação dos usuários
  - Aumentar a fidelidade dos usuários
  - Melhorar o gerenciamentos dos itens
- SRs estão sujeitos à falhas
  - Racismo?
- Problema está na consistência dos dados!

Research proves that consumer experience does matter

86%
CUSTOMERS

SAID THAT **PERSONALIZATION** HAS HAD SOME IMPACT ON PURCHASING DECISION

75%
SHOPPING TIME

IS SPENT ON PRODUCT **DISCOVERY & RESEARCH** ONLINE BY 50% CUSTOMERS

81%
CUSTOMERS

DEMAND IMPROVED RESPONSE TIME

95%
DATA

WITHIN ORGANIZATION REMAINS UNTAPPED

Source : http://www.nextopia.com/wp-content/uploads/2015/01/personalization-ecommerce-infographic.png
https://blog.hubspot.com/blog/tabid/6307/bid/23996/Half-of-Shoppers-Spend-75-of-Time-Conducting-Online-Research-Data.aspx
http://possible.mindtree.com/rs/574-LHH-431/images/Mindtree%20Shopper%20Survey%20Report.pdf
http://www.getelastic.com/using-big-data-for-big-personalization-infographic/

Métricas

- Identificar o melhor algoritmo de recomendação é um desafio
  - Discordância sobre os atributos e métricas
- Problemas ao avaliar os algoritmos:
  - Algoritmos dependem do conjunto de dados
  - Objetivos da avaliação podem variar

Algoritmos

- Utilizam como base informações e atributos de usuários e itens para recomendar que estão disponíveis no sistema. Estes são os principais componentes avaliados com base em diferentes critérios, cada qual com sua abordagem.

Netflix: Sistema de Recomendação

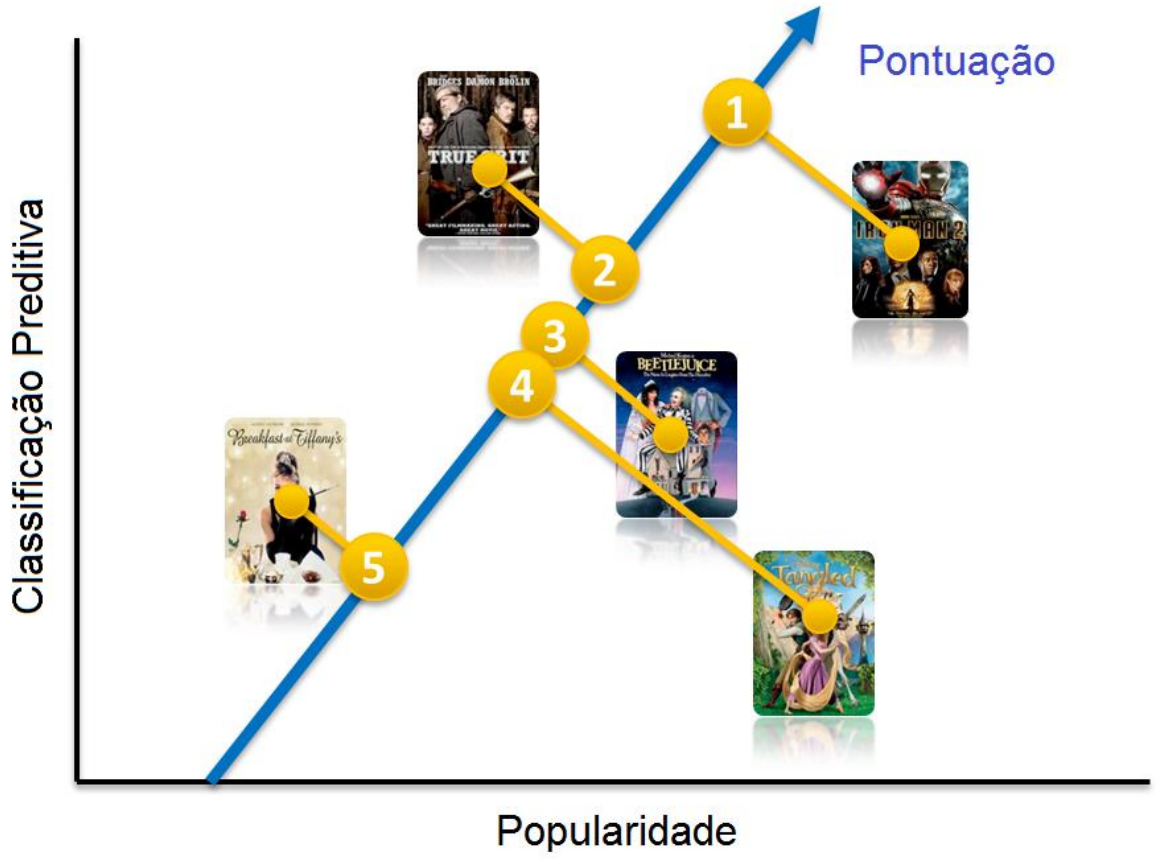
- Sistema de Recomendação Netflix
  - Tudo se torna uma recomendação!
  - Recomendações arranjadas em grupos colocados em linhas, e cada coluna é um item do grupo.

Netflix: Parâmetros Analisados

- Principais Parâmetros de Recomendação:
  - Semelhança (ou Similaridade);
  - Amigos (social);
  - Popularidade;
  - Gênero (ou Categoria);
  - Outros parâmetros podem incluir localização geográfica do usuário ou dados retirados de seu perfil ou outros acessos.
- Algoritmos observam estes parâmetros em conjunto, não separados.

Netflix: Ordenação de Itens

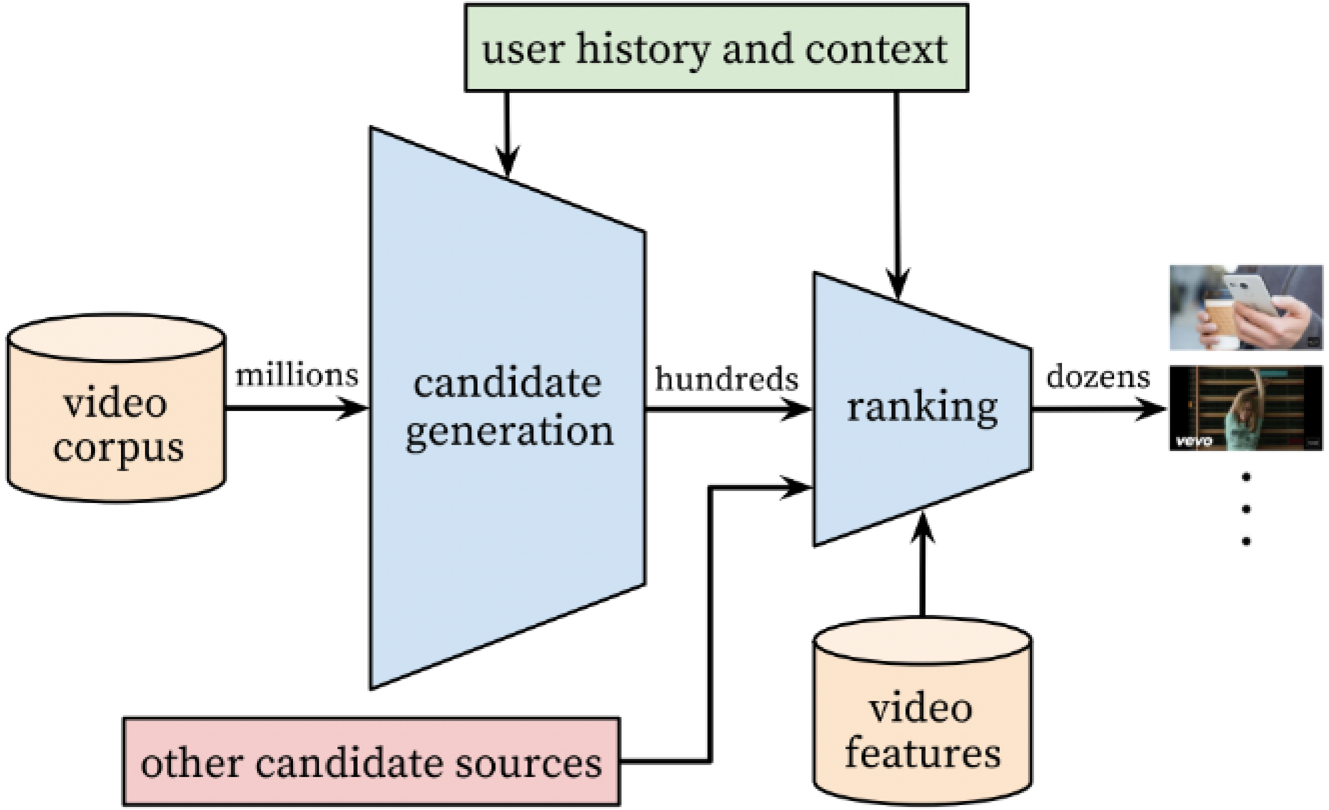
- Ordenação: exemplo simples de ordenação



Netflix: Conclusões

- Sistema de recomendação Netflix utilizada métodos híbridos;
- O sistema de recomendação é parte fundamental da experiência usuário – serviço, vez que boa parte da visualização de itens é por recomendação;
- A equipe Netflix busca constantemente a melhoria de seu sistema de recomendação direcionado a personalização, mostrando reconhecimento da importância deste como diferencial do serviço.

YouTube – arquitetura com 2 redes neurais



Deep Learning-based Search and Recommendation systems using TensorFlow

Abhishek Kumar
Dr. Vijay Agneeswaran

MARCH 06, 2018

SAPIENTRAZORFISH

Strata Conference – San Jose (2018)

YouTube – arquitetura com 2 redes neurais

- The **candidate generation** network takes the user's *activity history* (eg. IDs of videos being watched, search history, and user-level demographics) and outputs a few hundred videos that might broadly be applicable to the user.
The general idea is that this network should optimize for **precision;** each instance should be highly relevant, even if it requires forgoing some items which may be widely popular but irrelevant.

- In contrast, the **ranking** network takes a richer set of features for each video, and scoring each item from the candidate generation network. For this network, it's important to have a high **recall;** it's okay for some recommendation to not be very relevant as long as you're not missing the most relevant items**.**

ABHISHEK KUMAR

Senior Data Scientist , **SapientRazorfish**

Masters from **University of California, Berkeley**

Pluralsight Author

- Doing Data Science with Python
- R Programming Fundamentals
- Machine Learning with ENCOG
- Currently authoring : "*Deploying Machine Learning Models with Tensorflow Serving*"

DR. VIJAY AGNEESWARAN

Senior Director and Head of Data Science, **SapientRazorfish**

MS (Research) & PhD , **IIT Madras**

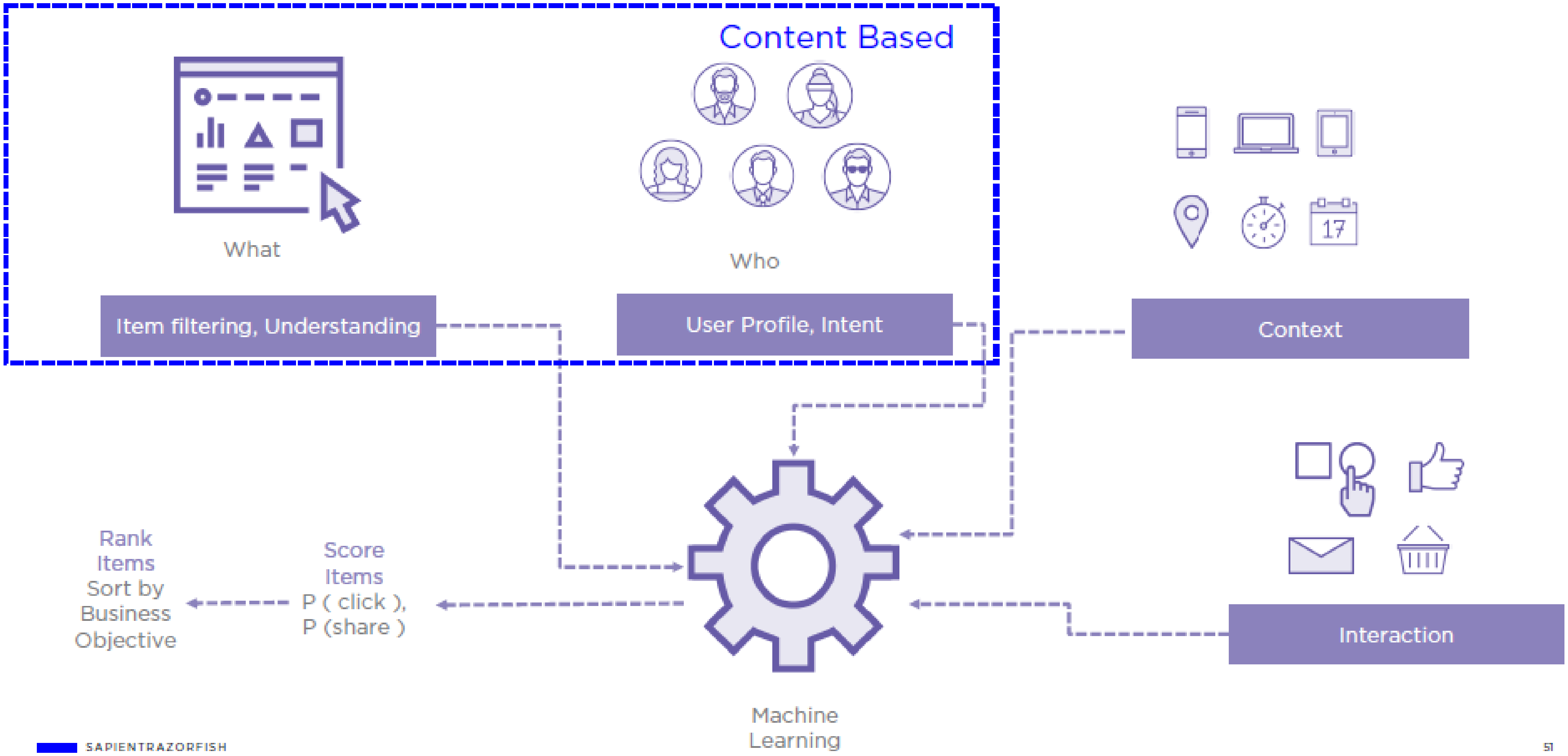
Post doctoral research fellowship, **LSIR Labs**

Professional member : **ACM, IEEE (Senior)**

4 Full US Patents and multiple publications (including **IEEE** journals)

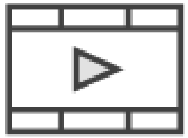
Regular Speaker @ **O'Reilly Strata conference**

Algoritmos


Content Based
What
Who
Item filtering, Understanding
User Profile, Intent
Context
17
Interaction
Rank
Items
Sort by
Business
Objective
Score
Items
P (click),
P (share)
Machine
Learning

RecSys 101 : Content Based Recommendation

Recommends an item to a user based upon a description of the item and a profile of the user's interests

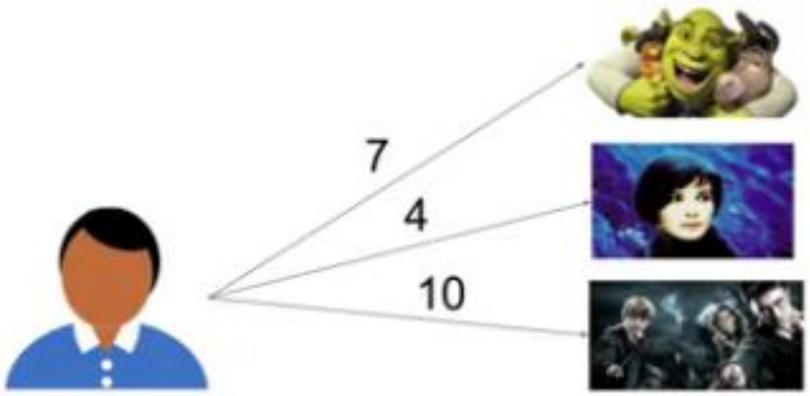


Representing Items using Features

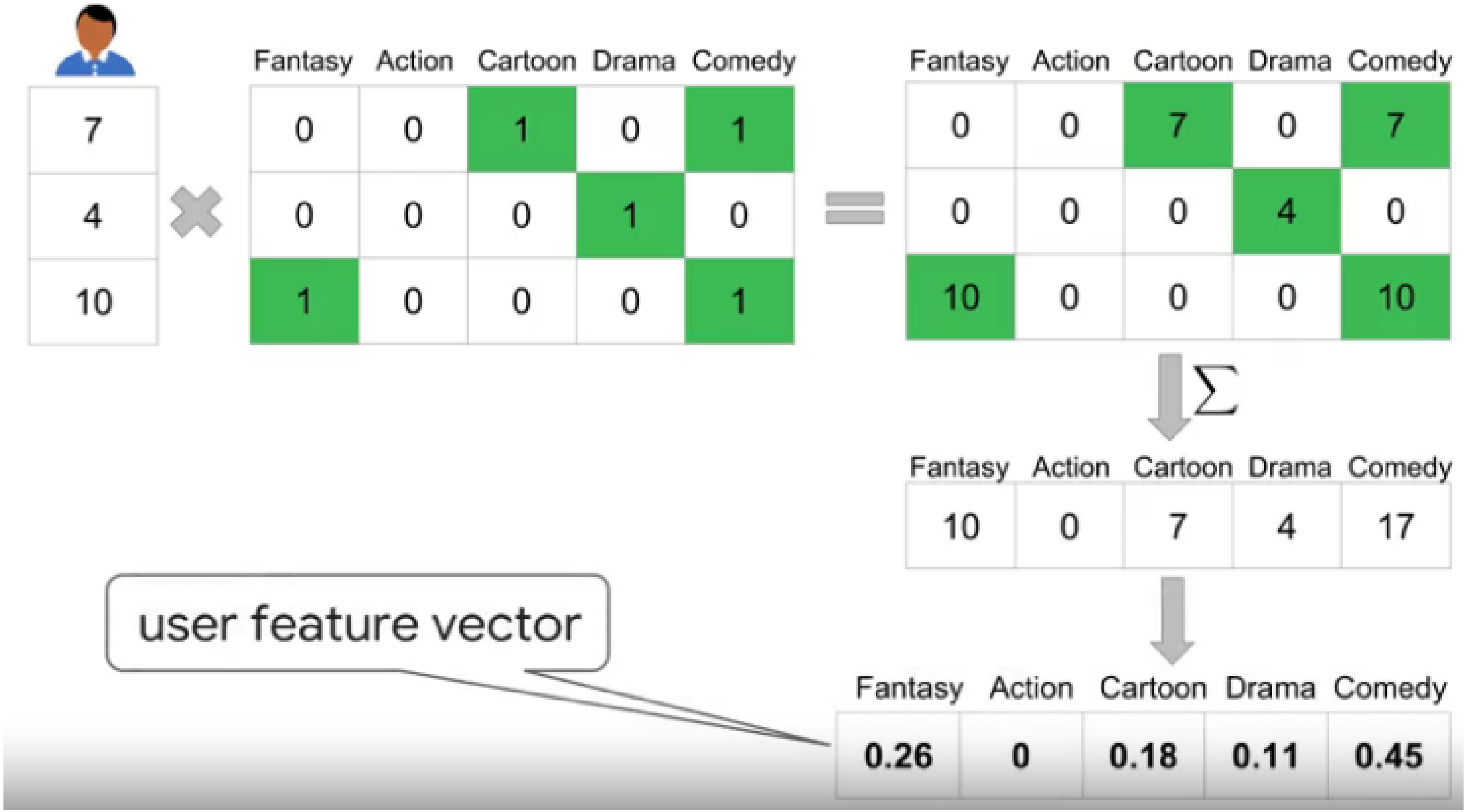
Drama	Arty	Comedy	Action	...	...	...	Commercial
0.7	0	0.2	0				0.8

User Profile

Creating a user profile that describes the types of items the user likes/dislikes

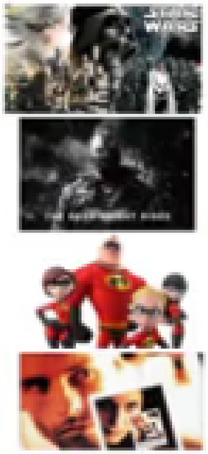


Fantasy	Action	Cartoon	Drama	Comedy
		1		1
			1	
1				1


Fantasy Action Cartoon Drama Comedy
7
4
10
0 0 1 0 1
0 0 0 1 0
1 0 0 0 1
Fantasy Action Cartoon Drama Comedy
0 0 7 0 7
0 0 0 4 0
10 0 0 0 10
Σ
Fantasy Action Cartoon Drama Comedy
10 0 7 4 17
user feature vector
Fantasy Action Cartoon Drama Comedy
0.26 0 0.18 0.11 0.45

0.26	0	0.18	0.11	0.45

×



1	1	0	0	1
1	1	0	1	0
0	1	1	0	1
0	0	0	1	0

=

0.26	0	0	0	0.45
0.26	0	0	0.11	0
0	0	0.18	0	0.45
0	0	0	0.11	0

$\xrightarrow{\Sigma}$

0.71
0.37
0.63
0.11

user movie ratings

RecSys 101 : Content Based Recommendation

Spotify

- More than 100 million monthly active users
- Over 30 million songs

Track: May 16

Artist: Lagawagon

Album: Let's Talk About Feelings

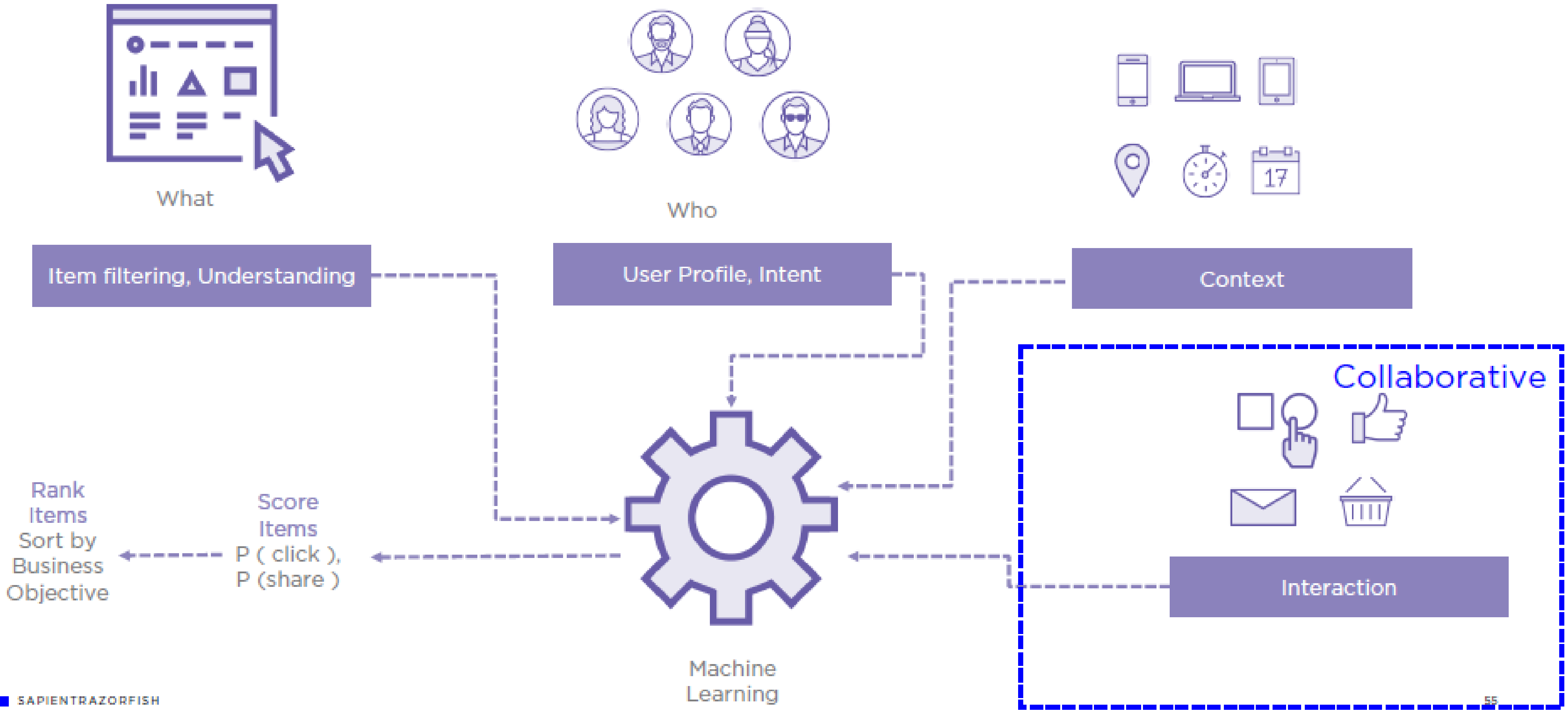
Release: 1998


```
{
  "danceability" : 0.560,
  "energy" : 0.527,
  "key" : 2,
  "loudness" : -9.783,
  "mode" : 1,
  "speechiness" : 0.0374,
  "acousticness" : 0.516,
  "instrumentalness" : 0.0000240,
  "liveness" : 0.156,
  "valence" : 0.336,
  "tempo" : 93.441,
  "type" : "audio_features",
  "id" : "2z7D7kbpRcTvEdT71tdiNQ",
  "uri" : "spotify:track:2z7D7kbp
  "track_href" : "https://api.spo
  "analysis_url" : "http://echone
  "duration_ms" : 168720,
  "time_signature" : 4
}
```

RecSys 101 : Content Based Recommendation

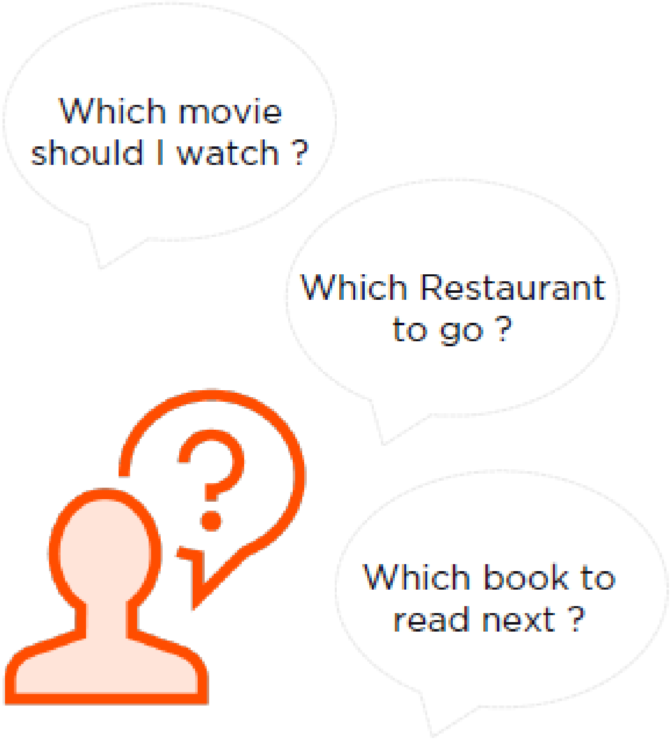
Pros	Cons
No need of other users data	Can only be effective in limited circumstances
Easy to understand reason behind recommendation	No suitable suggestions if content doesn't have enough information
Capable of recommending new and unknown items	Depend entirely on previous selected items and therefore cannot make predictions about future interests of users

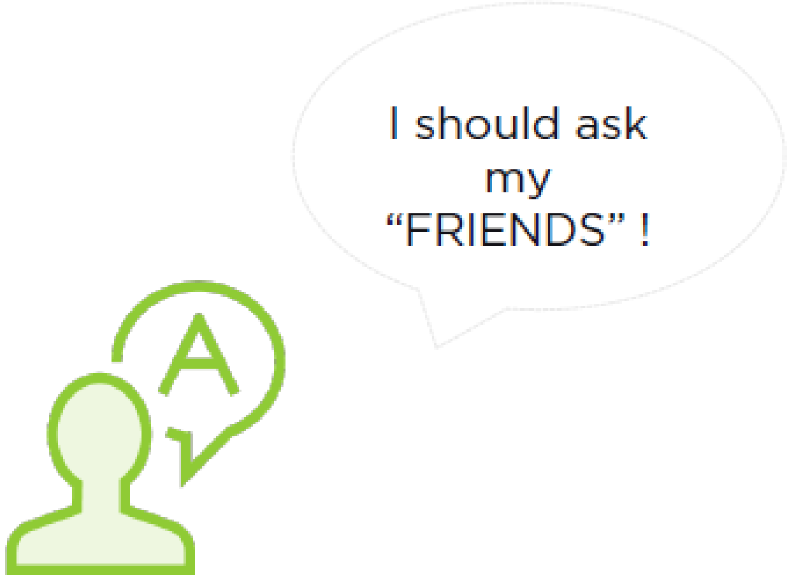
RecSys 101 : Internals



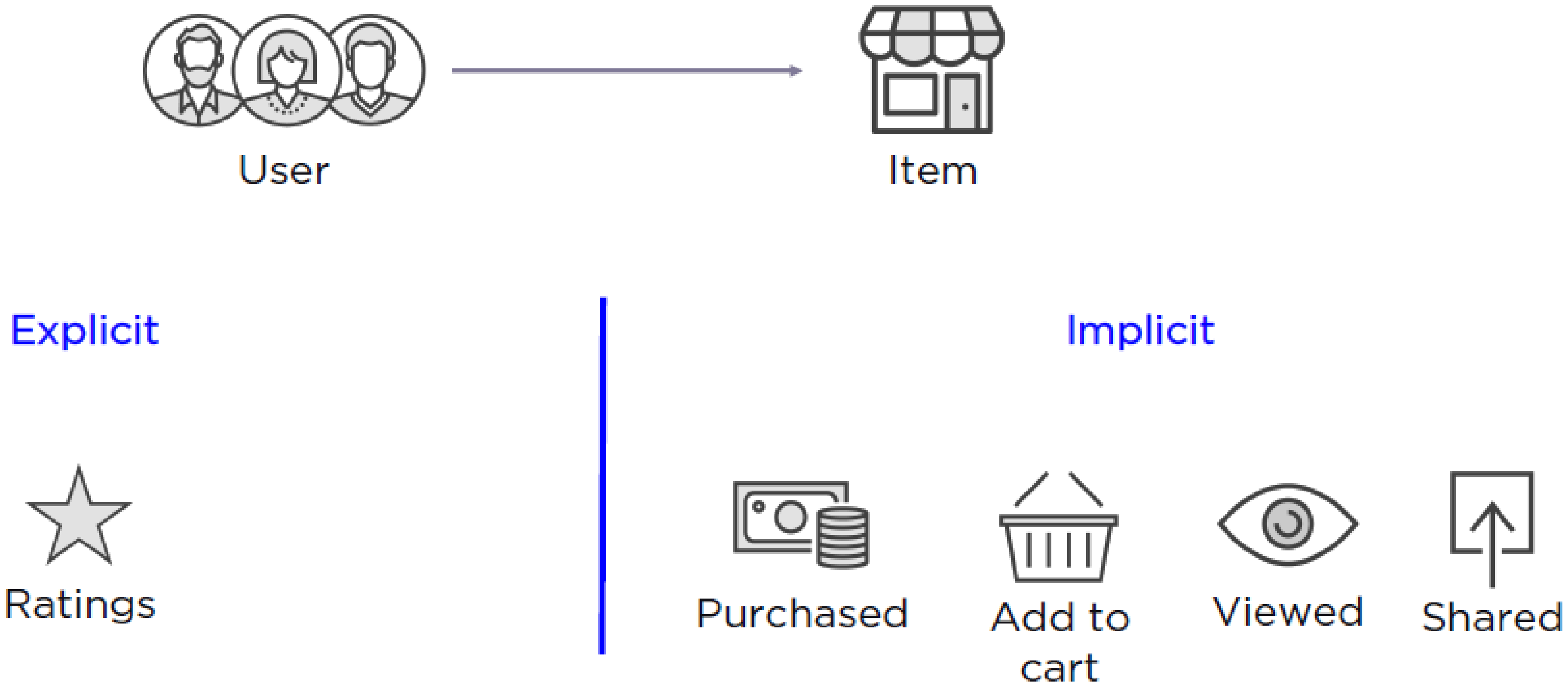
RecSys 101 : Collaborative Filtering

Unlike Content based filtering , Collaborative Filtering doesn’t require any product description at all





RecSys 101 : Collaborative Filtering : Interactions / Feedback



RecSys 101 : Collaborative Filtering : Interactions / Feedback

Explicit

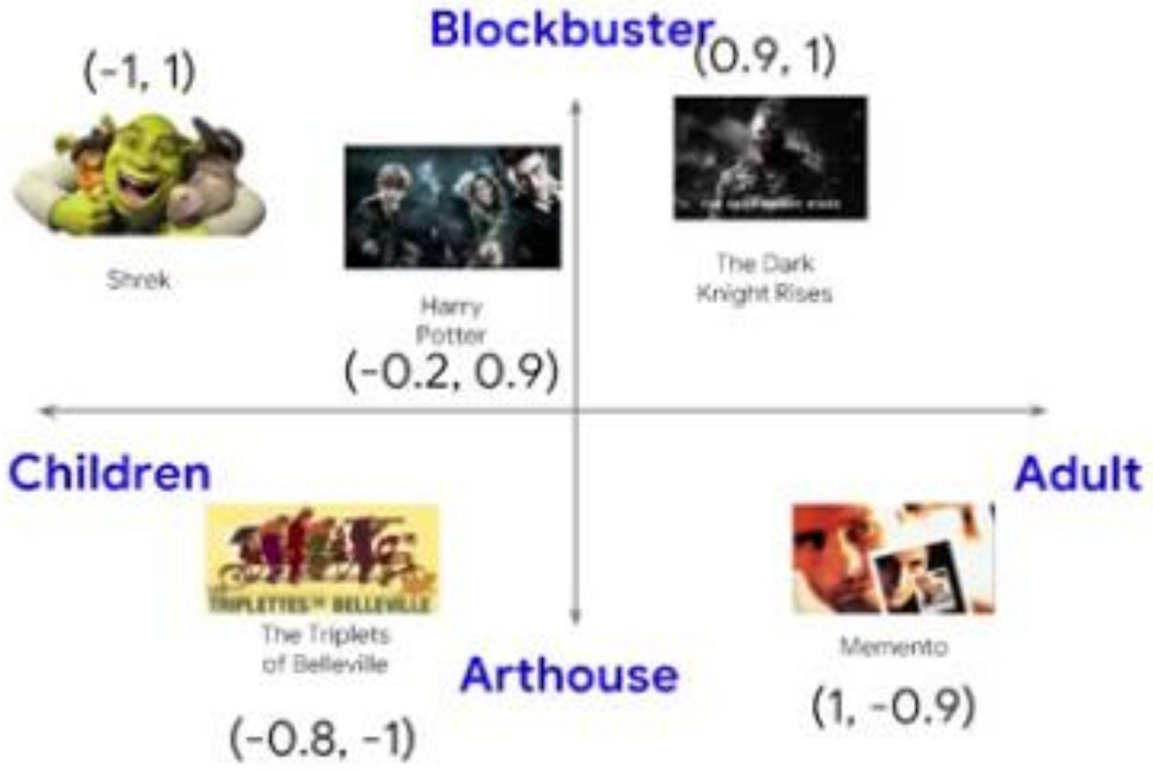
- Very few users leave ratings
- Very less explicit data
- Ratings are biased
- Often not easy for user to express likeness in terms of Ratings or score

Implicit

- Easy to track & Store web logs data
- Lots of implicit data generated for each user
- More the data , better the recommendations
- Noisy
- Difficult to infer Negative Feedback

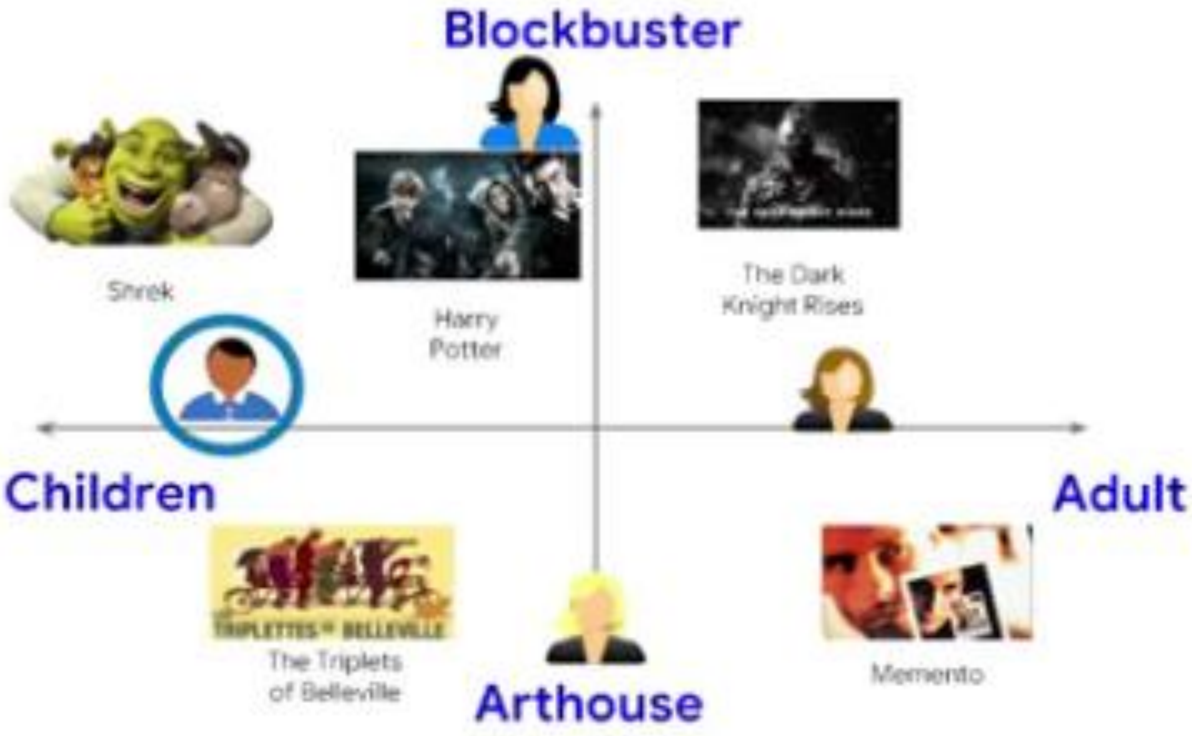
We can organize items by similarity in one dimension

We can organize items by similarity in two dimensions



Simply take the dot product between users and items in embedding space

Item retrieval in two dimensions based on user



Quiz

Based on this user-item interaction matrix, which movie should user 4 watch?

		0.9	-1	1	1	-0.9
		-0.2	-0.8	-1	0.9	1
1	0.1	✓		✓	✓	
-1	0		✓			✓
0.2	-1	✓	✓	✓		
0.1	1				✓	✓

A. Harry Potter
B. Triplets of Belleville
C. Shrek
D. The Dark Knight Rises

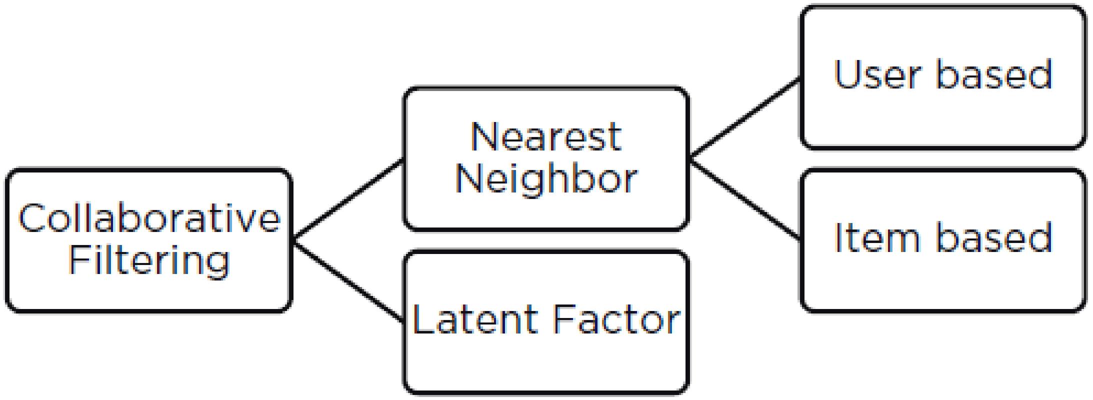
Quiz

Based on this user-item interaction matrix, which movie should user 4 watch?

			0.9	-1	1	1	-0.9
			-0.2	-0.8	-1	0.9	1
1	0.1		✓		✓	✓	
-1	0			✓			✓
0.2	-1		✓	✓	✓		
0.1	1		-0.11			✓	✓

A. Harry Potter **-0.11**
B. Triplets of Belleville **-0.9**
C. Shrek **-0.9**
D. The Dark Knight Rises **1.0**

RecSys 101 : Collaborative Filtering



→ Find similar users like me and recommend what they liked

→ Find similar items to those I have previously liked

Factor based techniques (Matrix Factorization, Factorization Machine)

$ 1 Million Netflix Challenge

- Scalability
- Predictive accuracy
- Can model real-life situations (e.g. Biases, Additional Input sources , Temporal Dynamics)

RecSys 101 : Collaborative Filtering : Latent Factor

Take the users and their feedback for different items and identify hidden factors that influence the user feedback

The idea is to factorize or decompose the user item matrix into two matrices

- Users are mapped on to hidden factors
- Items are mapped on to hidden factors

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	X		X		X	
User 2		X	X			
User 3				X		X
User 4					X	
User 5	X	X		X		X
User 6			X	X		
User 7	X	X	X		X	X
User 8		X		X		
User 9			X			

R

$\cong$

	UF1	UF2
User 1		
User 2		
User 3		
User 4		
User 5		
User 6		
User 7		
User 8		
User 9		

U

X

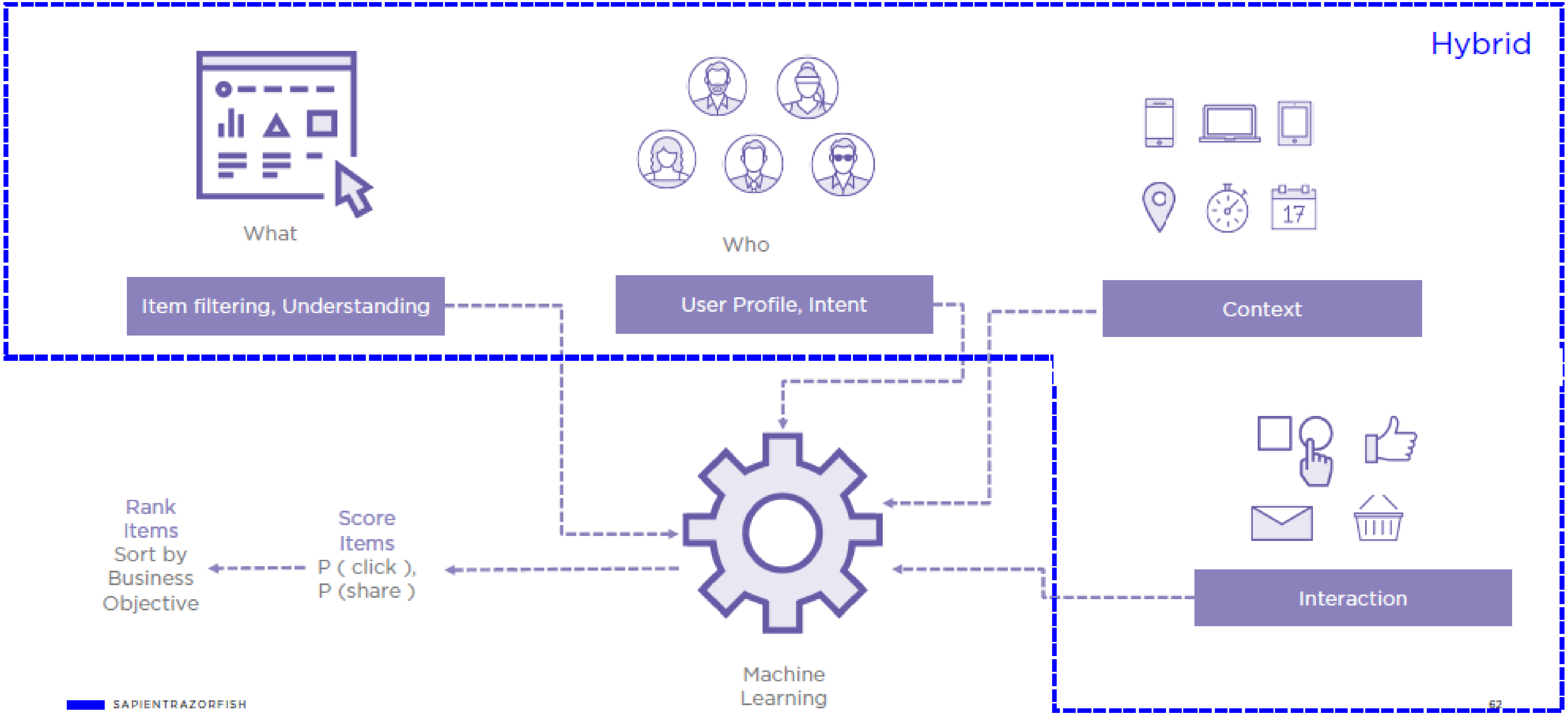
V

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
IF1						
IF2						

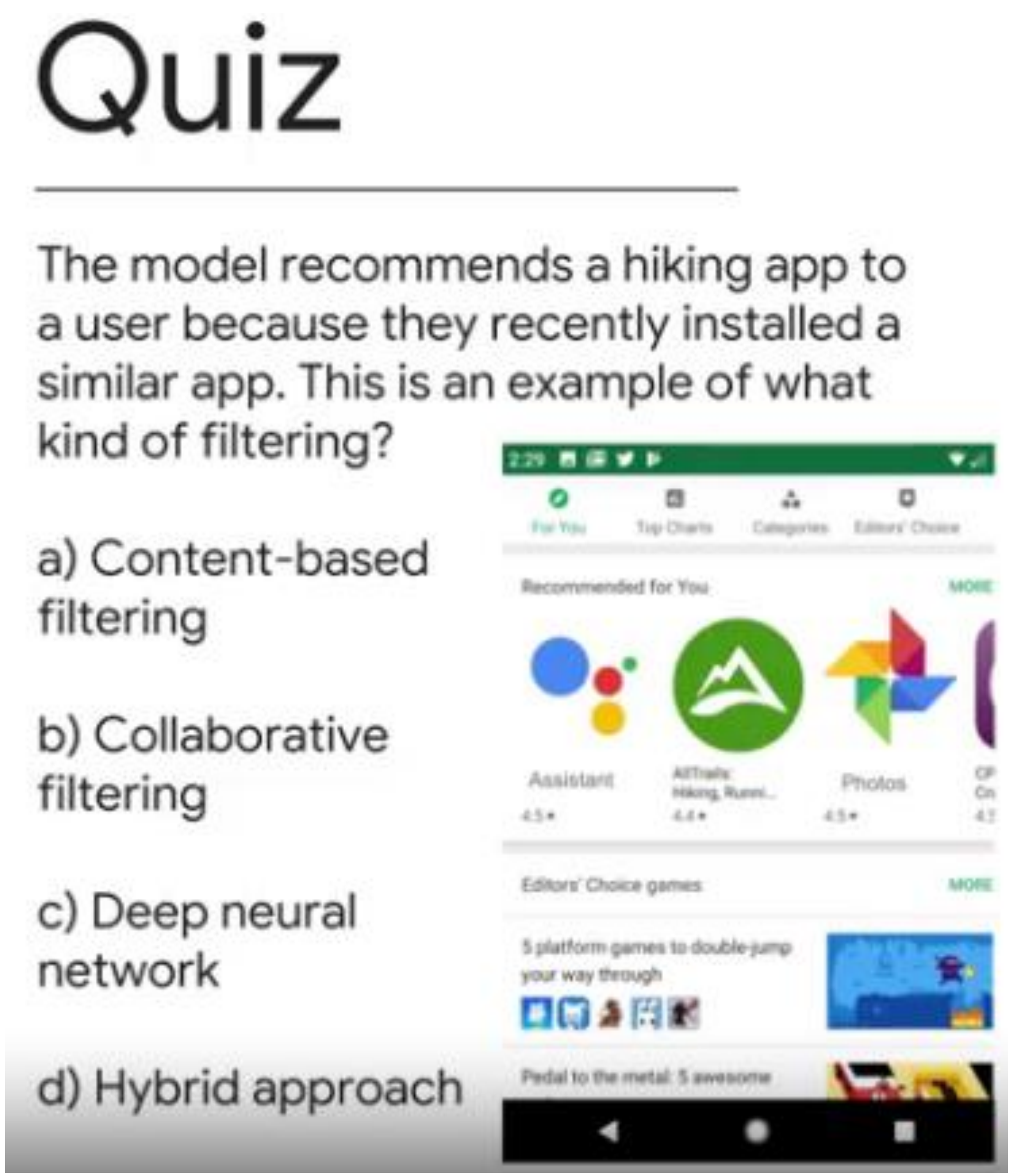
RecSys 101 : Collaborative Filtering

Pros	Cons
Content information not required either of users or items	Cannot produce recommendations if there is no interaction data available (Cold Start Problem)
Personalized recommendations using other user's experience	Often demonstrate poor accuracy when there is little data about users' ratings (Sparsity)
No domain experience required	Popular items get more feedback (Popularity bias)

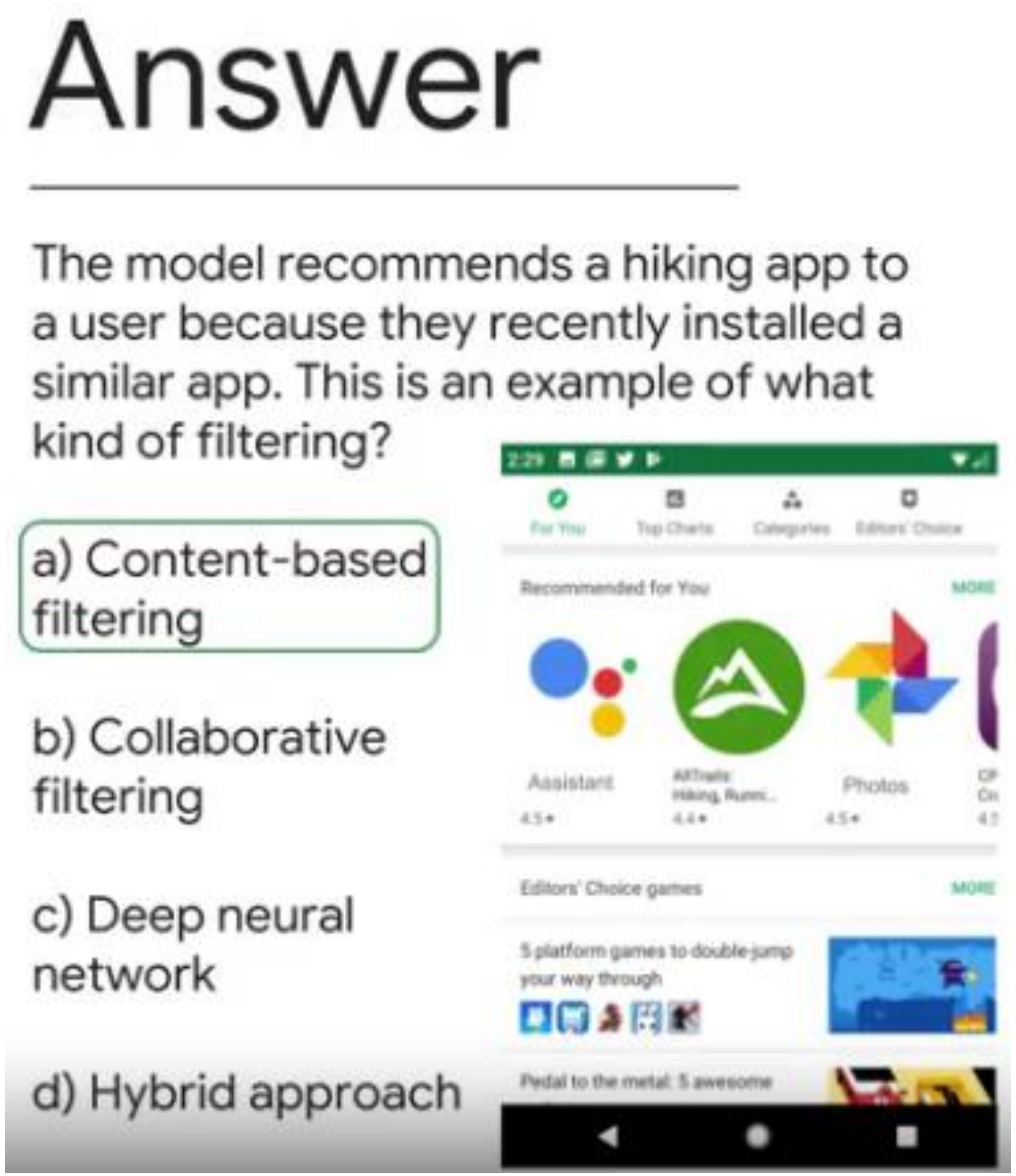
RecSys 101 : Internals



Quiz


Quiz
The model recommends a hiking app to a user because they recently installed a similar app. This is an example of what kind of filtering?
a) Content-based filtering
b) Collaborative filtering
c) Deep neural network
d) Hybrid approach
Recommended for You
MORE
Assistant
Photos
Editors' Choice games
MORE
5 platform games to double-jump your way through
Pedal to the metal: 5 awesome

Quiz


Answer
The model recommends a hiking app to a user because they recently installed a similar app. This is an example of what kind of filtering?
a) Content-based filtering
b) Collaborative filtering
c) Deep neural network
d) Hybrid approach
Recommended for You
MORE
Assistant
Photos
Editors' Choice games
MORE
5 platform games to double-jump your way through
Pedal to the metal: 5 awesome

RecSys 101 : Hybrid Recommendation Engine

Pros	Cons
Solve the issue of Cold Start by leverage both content and collaboration	Difficult to implement
Use of Implicit feedback reduces the sparsity issues to a large extent	
Can include higher order feature interactions as well	

Representation : A Key Aspect

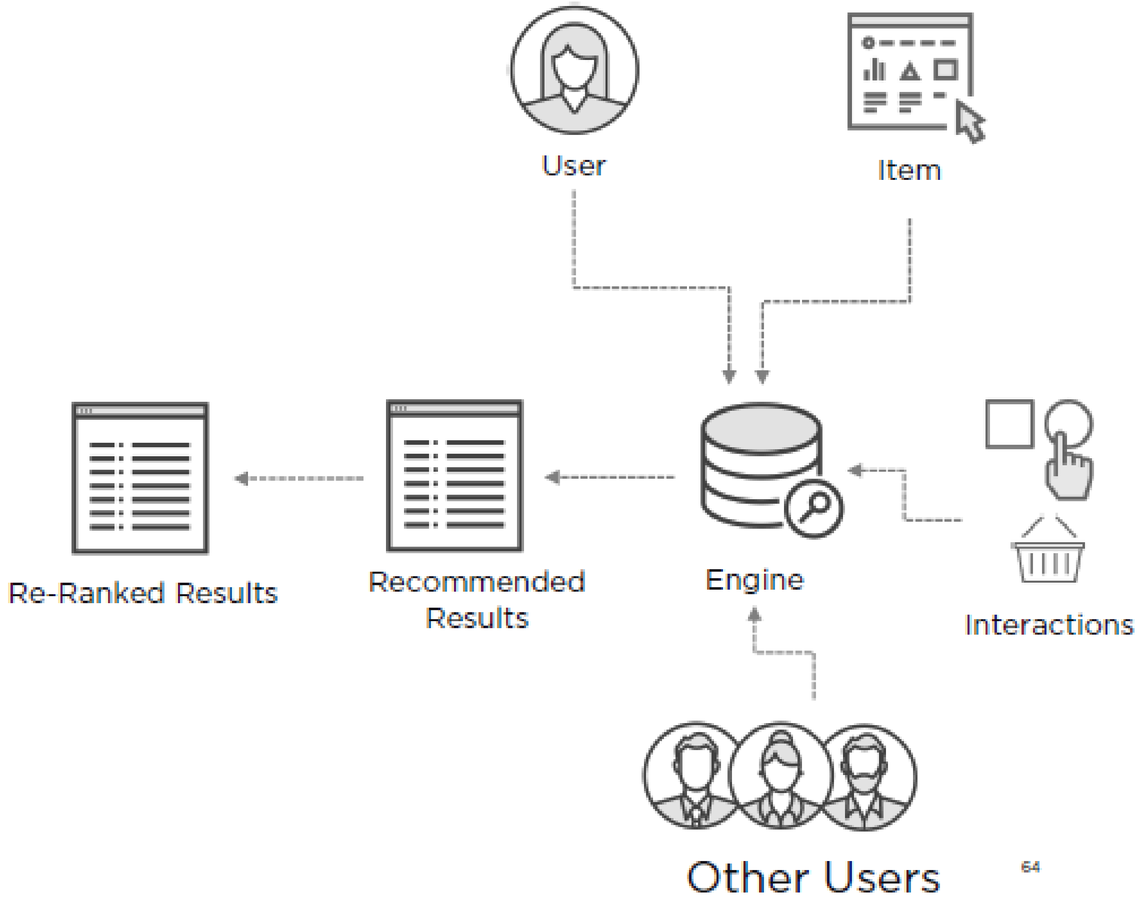
User ID

Item ID

User / Item Metadata

Representation

Recommendation Engines



Matrix Factorization

Items

Users

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	X		X		X	
User 2		X	X			
User 3				X		X
User 4					X	
User 5	X	X		X		X
User 6			X	X		
User 7	X	X	X		X	X
User 8		X		X		
User 9			X			

R

$\cong$

	UF1	UF2
User 1		
User 2		
User 3		
User 4		
User 5		
User 6		
User 7		
User 8		
User 9		

U

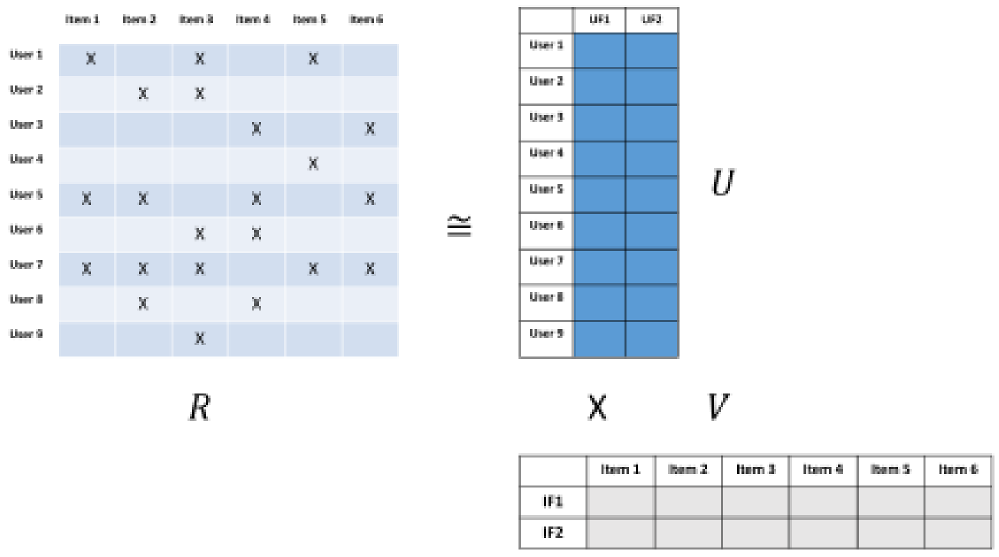
X V

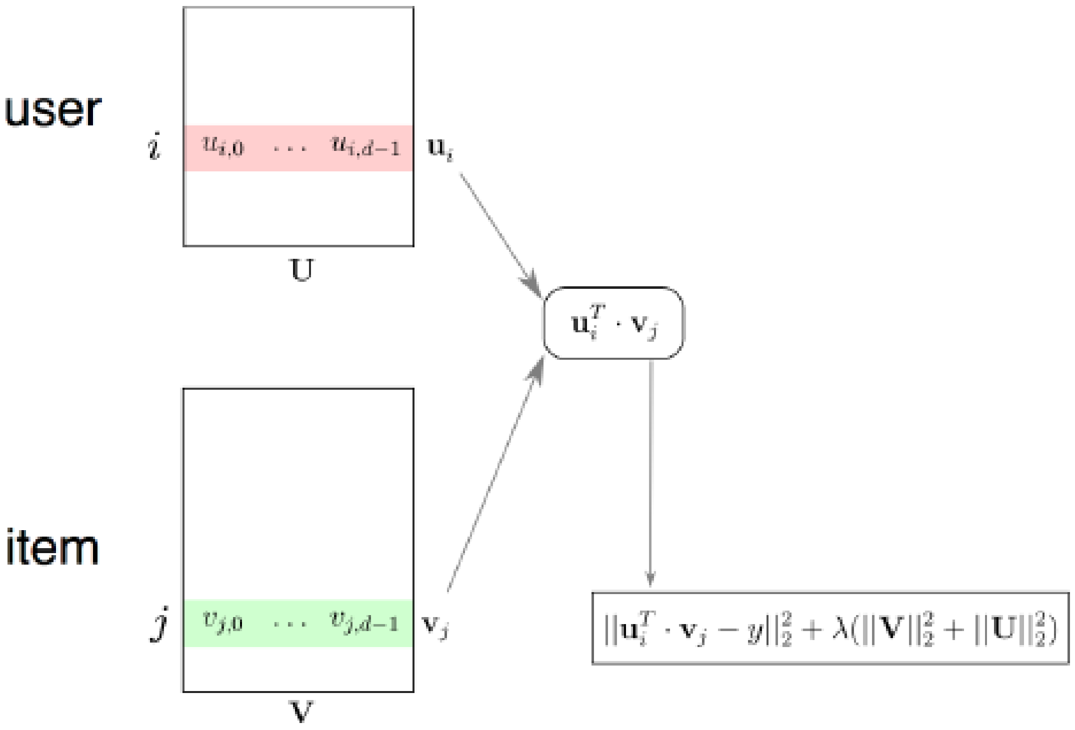
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
IF1						
IF2						

How to better represent users and items ?

What about item and user metadata ?

Matrix Factorization





Material adicional no Google

- https://cloud.google.com/solutions/machine-learning/recommendation-system-tensorflow-overview

Veja os tutoriais que acompanham esta visão geral:

- Criar o modelo (parte 1) mostra como usar o algoritmo WALS no TensorFlow para fazer previsões de classificação para o conhecido conjunto de dados MovieLens.
- Treinar e ajustar no Cloud Machine Learning Engine (parte 2) mostra como usar o Cloud Machine Learning Engine para treinar o modelo e empregar o recurso de ajuste de hiperparâmetro para otimizá-lo.
- Aplicar a dados do Google Analytics (parte 3) mostra como aplicar o sistema de recomendação a dados importados diretamente do Google Analytics 360 para realizar recomendações para sites que usam o Google Analytics.
- Implantar o sistema de recomendação (parte 4) mostra como implantar um sistema de produção no GCP para fazer recomendações em tempo real para um site.

Recomendação Híbrida com TensorFlow

Notebook ou vídeo “07-Recomendação Híbrida - lab”