Agenda

- Recap
- Introduction to Recommendation Sys
- Collaborative Filtering
  - User - User
  - Item - Item
- Cold Start problem
- Content Based Rec Sys
- RecSys as Classification / Regression

Introduction

Objective : Rec Sys suggests users relevant items based on their preference

Popular Apps :

- youtube / Netflix / Prime-video
- Spotify
- FB / insta
- Google News
- Zomato

History of RecSys:

Pre 2007 : Similarity Based Sys

Amazon

⇒ Collaborative

⇒ Content Based

2007 – 2015

Netflix → Hackathon / 1M $

⇒ Matrix Factorization

2015 → Ongoing

⇒ Deep learning Methods

* Data and Problem formulation

Users : 1 → n ; U_i (V Large) 100's mills

Items : 1 → m ; I_j (V large) 100's of millin

Task : $U_i \xrightarrow[\text{Recommend } K]{\text{Suggest}} (I_2, I_{10}, I_{20} \ldots)$

$\underbrace{\phantom{I_2, I_{10}, I_{20}}}_{\text{historical}}$ Pref

Data Representation

	I_1	I_2	I_3	------------	I_n
U_1	0	0	1		0
U_2	1	1	0		0
U_3	1	0	0		1
"					
"					
"					
U_n					

Sparse Matrix
(A matrix of 0 and 1's)

1 billion $\rightarrow$ yt Users

200 million $\Rightarrow$ yt Videos

Total Obs $\Rightarrow$ 1 bill $\times$ 200 mill

Sparsity Ratio: $\dfrac{\#\text{ non Empty Cells}}{\#\text{ Total Cells}}$

$\Rightarrow \dfrac{1\cancel{0}\cancel{0}}{100\cancel{0}\cancel{0}} \Rightarrow 0.01\% \Rightarrow$ 1 out 100 cells is filled

Collaborative Filtering

Sparse Matrix $\longrightarrow A_{n \times m}$

Users — Item

Task : $U_i \longrightarrow \text{items}_k$

$U_i \rightarrow$

I_1	I_2	I_3	- - - - - - - I_n
0	1	0	- - - - 1

$I_j \rightarrow$

U_1	U_2	U_3	- - - - - U_n
0	1	0	- - - - 1

$U_i \rightarrow$ Recommend some items

Idea 1 $\longrightarrow U_i \longrightarrow (i_2, i_n)$

0	1	0	- - - - 1

① all items which are similar to i_2

② all items which are similar to i_n

$i_2 \longrightarrow \boxed{Sim(i_2, item)_{1 \to N}}$ $K=3$

$\downarrow$

i_{10}, i_{12}, i_{15}

$i_n \longrightarrow \boxed{Sim(i_n, item_{1 \to N}}$ $K=3$

$\downarrow$

i_{30}, i_{18}, i_{12}

$U_i \longrightarrow I_{12}$

Item – Item Collaborative Filter

⊙ U_{12} liked

$\longrightarrow Sim(U_{12}, I_{1-m})$

Weekly Basis

Nightly

Sim $\boxed{\text{Item}_{1\to m} \,,\, \text{Item}_{1\to m}}$

↓

Store

	I_1	I_2	I_3	---- m
I_1	1	0.8	0.3	
I_2	0.8	1	0.9	
I_3	0.3	0.9	1	
⋮ m				

new ↓

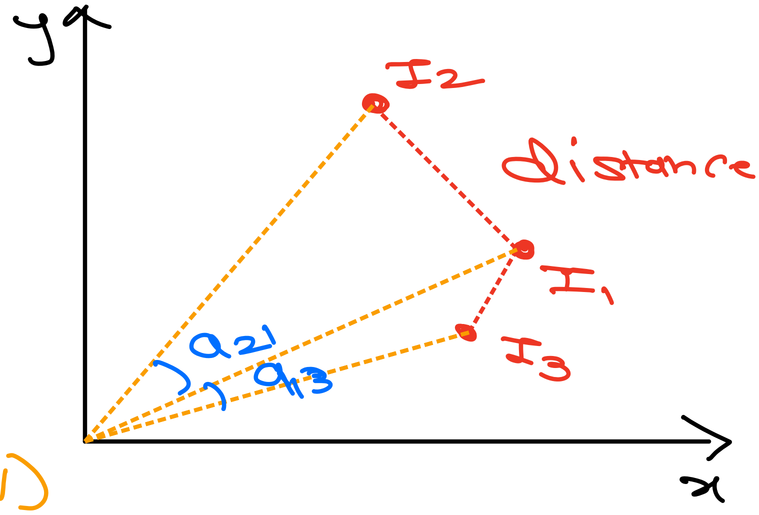
Similarity Matrix

Caching

Similarity Function

1) Distance
  ↳ Euclidean
  ↳ Manhattan
  99

2) Cosine Similarity
$(U \cdot V) / |U||V|$



$I_1 \Rightarrow [0, 1, 1, 0, 1, 1, 0 - \dots]$

$I_9 \Rightarrow [0, 1, 1, 0, 1, 1, 1, 1 - \dots]$

$\boxed{\text{User – User CF}}$

$U_i \longrightarrow$ Similar Users

$\downarrow$

Most Common Items

$Sim(U_i, U_k) \Rightarrow cos\_sim(U_i, U_k)$

$U_i \longrightarrow I_{10}, I_{18}$

$U_{10} \Rightarrow I_{12}, I_{10}, I_{18}, I_{20}$

$U_{269} \; I_{18}, I_{10}, I_{56}$

$U_{589} \; I_{10}, I_{18}, I_{20}$

$U_i \xrightarrow{\text{Suggest}} 20, 56, 12$

User — User Interaction

Collaboration

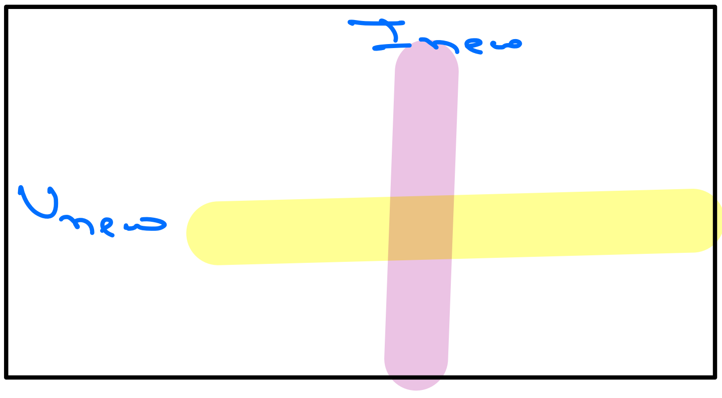
$\boxed{\text{Cold-Start}}$

$U_{new} \Rightarrow$

0	0	0	0 0 0 0

$I_{new} \Rightarrow$

0	0	0	0 0 0 0



Sparse Matric

Content Based

→ Top 10 most seen, Most liked frequently accessed Items (generic)

* Personalisation based on Metadata

Age:

location: Track from ip
GPS

Type of Credit_card:

Device: Android, Mac

Tag's:

Language:

	f_1	f_2	f_3	- - - - - - - -
$U_i \Rightarrow$				
$U_k \Rightarrow$				

$I_k \rightarrow$

f_1	f_2	f_3	- - -

U_i	f_1	f_1	- - - - -
U_1			
U_2			

Feature Matrix

Disadvantage :

* Manully Engineered features

Time and Domain

Advantage

* No Cold start issue