Overview of Recommender Systems

Recommendations are Everywhere!

Popular recommenders:

- Family and Friends
- Colleagues
- Internet

Recommendation System (RecSys)

Recommendation System is an information filtering technique, that provides users with information he/she may be interested in.

66% Rented Movies from recommendations

35% Sales from recommendations

Source: Celma & Lamere, ISMIR 2007

Popular Approaches in RecSys

- Rule-based
- Content-based filtering
- Collaborative filtering
- Hybrid Recommender Systems

Content Based Filtering

What it is?

- Recommends items based on comparison between the content of the item and user profile
- Content of an item is represented by set of descriptors or features
- User profile shows “affinity” of user towards same descriptors, derived using items user has “liked” in past
- For new item, its features are calculated and then compared with user profile to decide the recommendation

Popular Descriptors: Keywords, TF-IDF, word2vec

Content based Movie Recommendation

Action	1	1	0	1
Romance	0	0	1	0

Content based Movie Recommendation

Action Affinity	Romance Affinity
0.7	0.3

Content based Movie Recommender

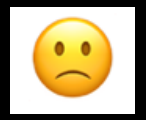
?

?

Action	Romance
0.7	0.3

Action	Romance
1	0

Action	Romance
0	1



Advantages & Disadvantages of Content RecSys

Advantages:

- User Independence
- Transparency
- New Item

Disadvantages:

- Limited Content analysis
- Over Specialization

Collaborative Filtering

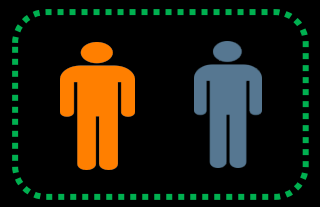
What is Collaborative Filtering?

- Recommend an item to a user based on preferences or taste of many similar users
- Key assumption is people who agreed with you in the past might do so in future

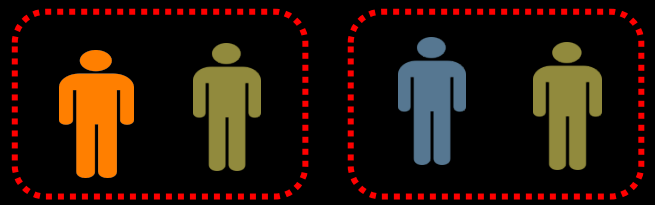
Implementation Strategies

- User Based
- Item Based
- Matrix Factorization

User Based

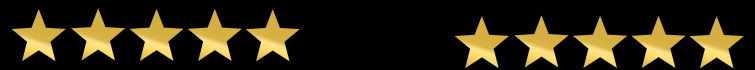


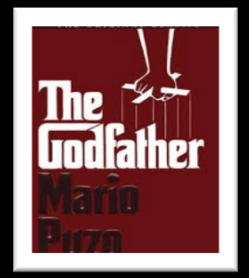
Similar Viewers
(Nearest Neighbors)

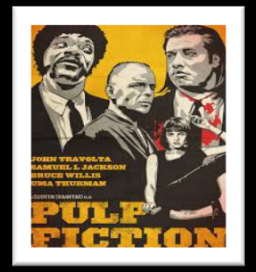


Dissimilar Viewers

Item- Item Based







Similar Movies

Dissimilar Movies

Matrix Factorization (SVD)

Movies from 1 to M

Users from 1 to N

	M_1	M_2	M_{m-1}	M_m
U_1				
U_2				
U_{n-1}				
U_n				

n x m matrix

Utility Matrix

=

Latent Features from 1 to K

Users from 1 to N

	K_1	K_{k-1}	K_k
U_1			
U_2			
U_{n-1}			
U_n			

n x k matrix

X

Latent Features from 1 to K

Latent Features from 1 to K

k x k diagonal matrix

X

Movies from 1 to M

Latent Features from 1 to K

	M_1	M_2	M_{m-1}	M_m
K_1				
K_{k-1}				
K_k				

k x m matrix

Matrix Factorization - Example

	Godfather	Pulp Fiction	Donnie Darko	The Notebook
U_1	5	4	4	2

=

	Thriller	Romantic	Scifi
U_1	1	0	1

X

Thriller	5		
Romantic		1	
Scifi			4

X

	Godfather	Pulp Fiction	Donnie Darko	The Notebook
Thriller	4	5	4	1
Romantic	1	2	3	5
Scifi	1	2	4	1

Collaborative Filtering Over Content Filtering

Advantages:

- CF doesn’t require item content analysis or extraction
- Can capture subtle things i.e. Diverse and Serendipity recommendations: via latent factors

Disadvantages:

- Sparsity of user preferences: hard to find users with similar taste
- Item cold start
- Gray Sheep and Shattacks

Hybrid Recommender Systems

Combination of Content based and Collaborative filtering!

Can include demographics and other knowledge also.

Some ways to achieve hybridization:

- Weighted
- Switching
- Mixed
- Cascade

Evaluation Methods & Metrics

Online Evaluation:

- A/B Testing to monitor change in CTR/CR

Offline Evaluation:

- RMSE/MAE for explicit feedback
- Recall for implicit feedback