

Movie Recommender System

Carol Duan

What is a recommender system?

A system that is able to provide or suggest items to the end users



... long live the Age of Recommendation!

“We are leaving the age of information
and entering the age of recommendation”

-- Chris Anderson in “The Long Tail”

Almost every major tech company has applied recommender system in some form or the other...



Amazon: 35% sales from recommendations

PEOPLE YOU MAY KNOW



Netflix: 2/3 of the movies watched are recommended

NETFLIX



Adventure Time



Monster Math Squad



Pajanimals



Winnie the Pooh



Julius Jr.



amazon.com



Who to follow · Refresh · View all



Joyent @joyent

Followed by Park Hoon and ...

Follow Promoted



APIdays Global @APIdaysG...

Follow

Find friends



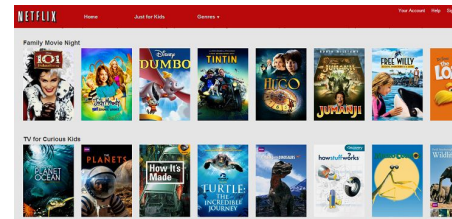
A Movie Recommender System

Dataset:

MovieLens 1M anonymous ratings of approximately 3,900 movies made by 6,040 MovieLens users who joined MovieLens in 2000.

Goal: Build a movie recommender system

- For new user, recommend new similar movies by searching movie titles
- For existing user, predict their ratings for the unseen movies and give recommendations



Movie features

	MovieID	Title	Genres
0	1	Toy Story (1995)	Animation Children's Comedy
1	2	Jumanji (1995)	Adventure Children's Fantasy
2	3	Grumpier Old Men (1995)	Comedy Romance
3	4	Waiting to Exhale (1995)	Comedy Drama
4	5	Father of the Bride Part II (1995)	Comedy

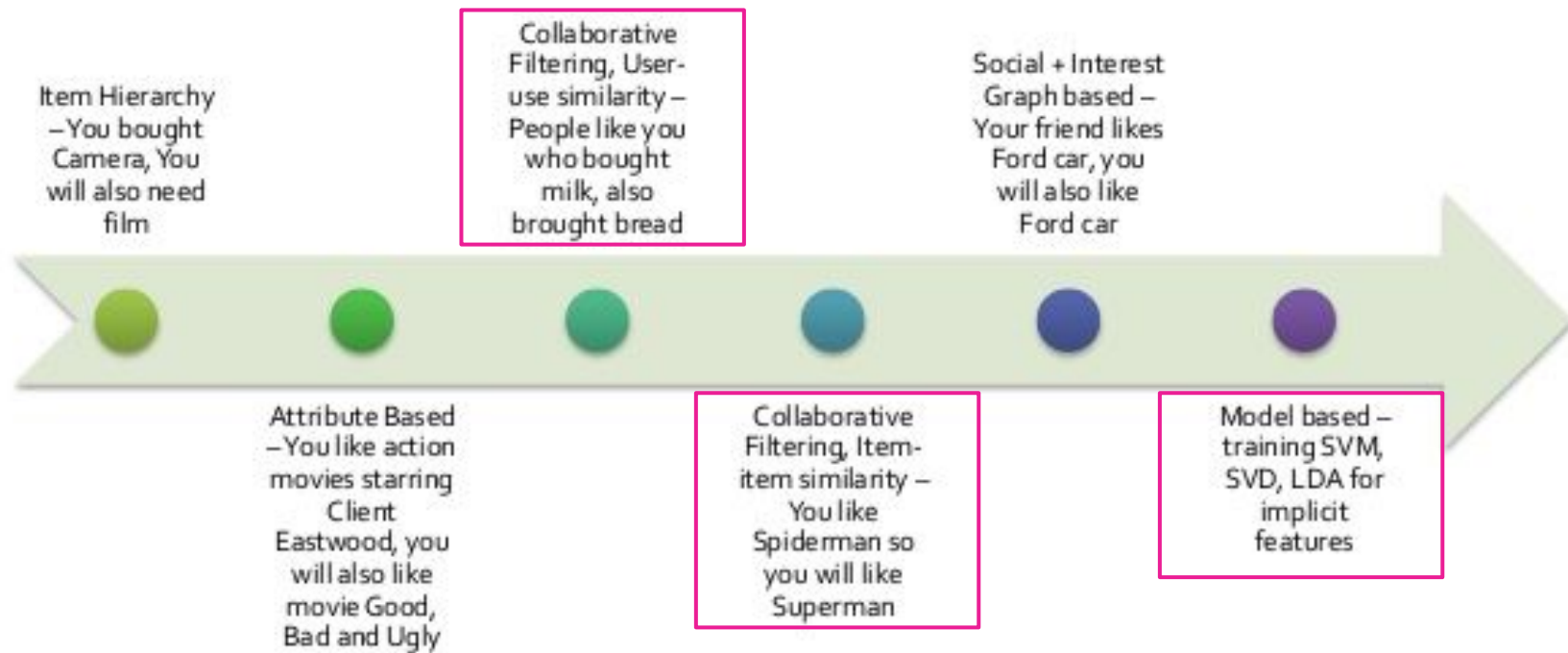
User features

	UserID	Gender	Age	JobID
0	1	F	1	10
1	2	M	56	16
2	3	M	25	15
3	4	M	45	7
4	5	M	25	20

Existing movie ratings

	UserID	MovieID	Rating
0	1	1193	5
1	1	661	3
2	1	914	3
3	1	3408	4
4	1	2355	5

Evolution of recommender system...



Collaborative / content-based recommender system

Predict ratings based on ratings from / of similar users or movies

UserID	MovieID	Rating
0	1	1193
1	1	661
2	1	914
3	1	3408
4	1	2355

INPUT

4	3			5	
5		4		4	
4		5	3	4	
	3				5
	4				4
		2	4		5

Introduction to Recommender Systems
Machine learning Paradigms
Social Network-based Recommender Systems
Learning Spark
Recommender Systems Handbook
Recommender Systems and the Social Web

USER-BASED COLLABORATIVE FILTERING

	1.00	0.75	0.63	0.22	0.30	0.00
	0.75	1.00	0.91	0.00	0.00	0.16
	0.63	0.91	1.00	0.00	0.00	0.40
	0.22	0.00	0.00	1.00	0.97	0.64
	0.30	0.00	0.00	0.97	1.00	0.53
	0.00	0.16	0.40	0.64	0.53	1.00

$$(0.7 \times \text{Movie 1}) + (0.6 \times \text{Movie 2}) = \text{Movie 1} \quad \text{already rated by user}$$

$$(0.7 \times 4 + 0.6 \times 5) / (0.7 + 0.6) = 3.0$$

$$(0.6 \times 3) / 0.6 = 3.0$$

$$= 3.0$$

How about other features?

ITEM-BASED COLLABORATIVE FILTERING

	1.00	0.27	0.79	0.32	0.98	0.00
	0.27	1.00	0.00	0.00	0.34	0.65
	0.79	0.00	1.00	0.69	0.71	0.18
	0.32	0.00	0.69	1.00	0.32	0.49
	0.98	0.34	0.71	0.32	1.00	0.00
	0.00	0.65	0.18	0.49	0.00	1.00

$$(4 \times \text{Movie 1}) + (3 \times \text{Movie 2}) + (5 \times \text{Movie 3}) = \text{Movie 1} \quad \text{already rated by user}$$

$$(0.8 \times 4 + 0.7 \times 5) / (0.8 + 0.7) = 4.5$$

$$(0.7 \times 3) / 0.7 = 3.0$$

$$= 3.0$$

CONTENT-BASED FILTERING

	1.00	0.00	0.58	0.00	0.67	0.58
	0.00	1.00	0.00	0.41	0.00	0.00
	0.58	0.00	1.00	0.00	0.58	0.75
	0.00	0.41	0.00	1.00	0.00	0.00
	0.67	0.00	0.58	0.00	1.00	0.58
	0.58	0.00	0.75	0.00	0.58	1.00

$$(4 \times \text{Movie 1}) + (3 \times \text{Movie 2}) + (5 \times \text{Movie 3}) = \text{Movie 1} \quad \text{already rated by user}$$

$$(0.4 \times 3) / 0.4 = 3.0$$

$$(0.6 \times 4 + 0.6 \times 5) / (0.6 + 0.6) = 4.5$$

$$= 3.0$$

OUTPUTS



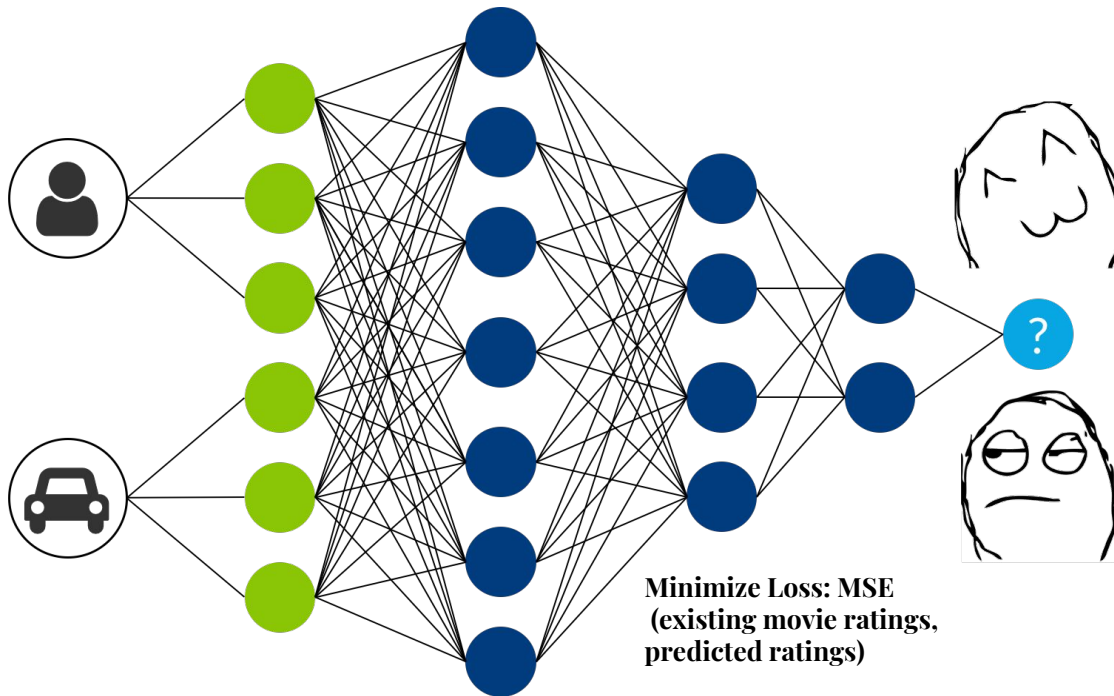
Model-based recommender system (Neural Network)

User features

	UserID	Gender	Age	JobID
0	1	F	1	10
1	2	M	56	16
2	3	M	25	15
3	4	M	45	7
4	5	M	25	20

Movie features

	MovieID	Title	Genres
0	1	Toy Story (1995)	Animation Children's Comedy
1	2	Jumanji (1995)	Adventure Children's Fantasy
2	3	Grumpier Old Men (1995)	Comedy Romance
3	4	Waiting to Exhale (1995)	Comedy Drama
4	5	Father of the Bride Part II (1995)	Comedy



Predicted ratings

	UserID	MovieID	Rating
0	1	1193	5
1	1	661	3
2	1	914	3
3	1	3408	4
4	1	2355	5

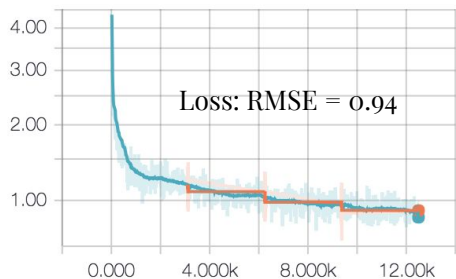
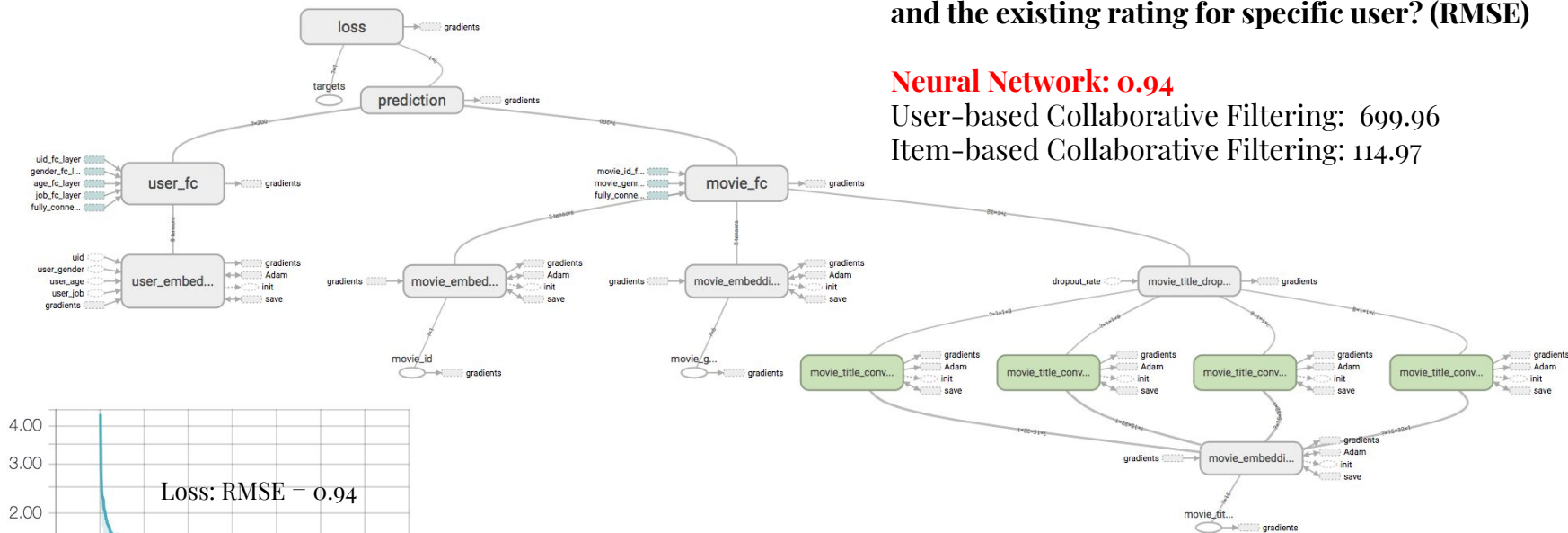
Model Implementation and Evaluation (TensorFlow)

How different between the movie rating we predict and the existing rating for specific user? (RMSE)

Neural Network: 0.94

User-based Collaborative Filtering: 699.96

Item-based Collaborative Filtering: 114.97



Showcase: New Users

Recommend similar new movies by searching movie titles

The movie you searched:

Movie ID: 1210
Movie Title: Star Wars: Episode VI - Return of the Jedi (1983)
Movie Genres: Action|Adventure|Romance|Sci-Fi|War
Average Rating: 4.0

Here are five movies you may like:

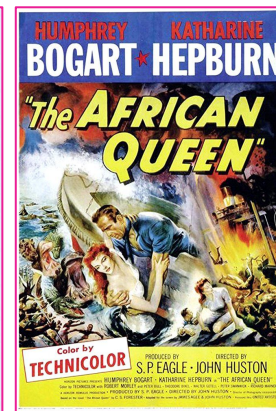
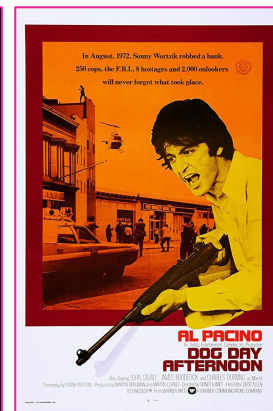
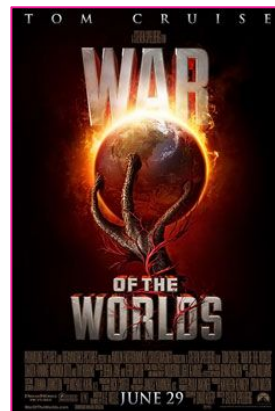
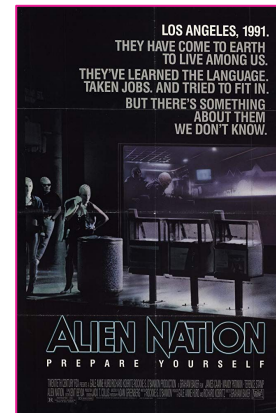
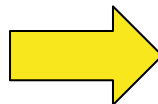
Movie ID: 2662
Movie Title: War of the Worlds, The (1953)
Movie Genres: Action|Sci-Fi|War
Average Rating: 3.9

Movie ID: 1372
Movie Title: Star Trek VI: The Undiscovered Country (1991)
Movie Genres: Action|Adventure|Sci-Fi
Average Rating: 3.4

Movie ID: 3701
Movie Title: Alien Nation (1988)
Movie Genres: Crime|Drama|Sci-Fi
Average Rating: 3.2

Movie ID: 3362
Movie Title: Dog Day Afternoon (1975)
Movie Genres: Comedy|Crime|Drama
Average Rating: 4.0

Movie ID: 969
Movie Title: African Queen, The (1951)
Movie Genres: Action|Adventure|Romance|War
Average Rating: 4.3



Showcase: Existing Users

Recommend new movies you may like based on your profile (user info / movie ratings)

Here is your profile:

User ID: 4
User Gender: M
User Age: 45
Average Rating: 4.2

Here are five movies you may like:

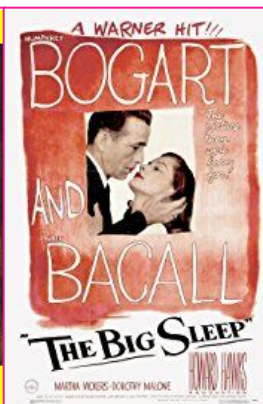
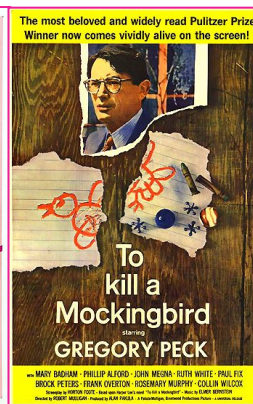
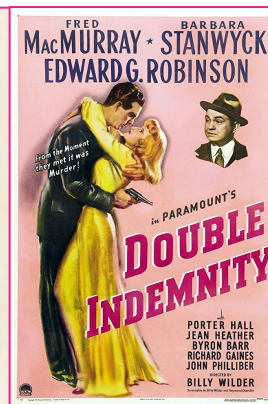
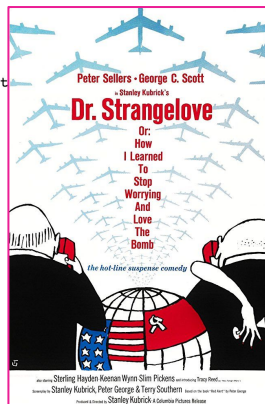
Movie ID: 750
Movie Title: Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb
Movie Genres: Sci-Fi|War
Average Rating: 4.4

Movie ID: 913
Movie Title: Maltese Falcon, The (1941)
Movie Genres: Film-Noir|Mystery
Average Rating: 4.4

Movie ID: 3435
Movie Title: Double Indemnity (1944)
Movie Genres: Crime|Film-Noir
Average Rating: 4.4

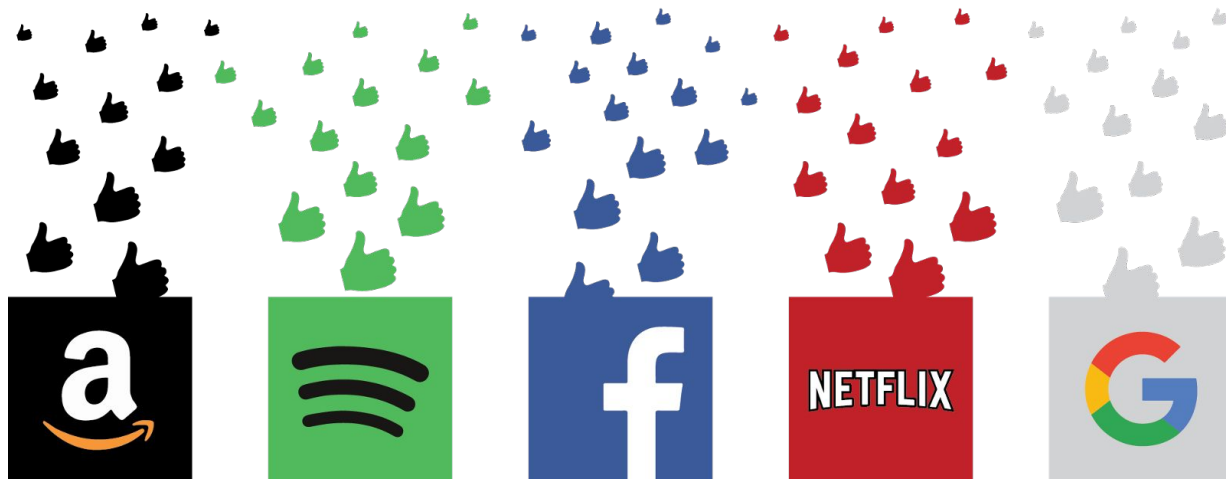
Movie ID: 1207
Movie Title: To Kill a Mockingbird (1962)
Movie Genres: Drama
Average Rating: 4.4

Movie ID: 1284
Movie Title: Big Sleep, The (1946)
Movie Genres: Film-Noir|Mystery
Average Rating: 4.3



Next Step

1. Train the model on a bigger dataset, add more movies
2. Explore other models (e.g. SVD) and compare the score
3. Build an user interface for better interactions



Thanks for Watching!

