



RECOMMENDER SYSTEMS

MOVIES

Andreea Stroia
Hala Albahloul

Today's Discussion

Topics to Cover

- Content Based Filtering
- Collaborative Filtering
- Hybrid





Collaborative Filtering

system matches persons with similar interests and provides recommendations based on this matching

	userId	movieId	rating	timestamp
0	1	31	2.5	1260759144
1	1	1029	3.0	1260759179
2	1	1061	3.0	1260759182
3	1	1129	2.0	1260759185
4	1	1172	4.0	1260759205

Basic approach

ITEM TO ITEM

Building a pivot table

movieId	1	2	3	4	5	6	7	9	10	11	...	119145	122882	122886	122892
userId															
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	...	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	...	0.0	0.0	0.0	0.0
5	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
...
667	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	3.0	...	0.0	0.0	0.0	0.0
668	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
669	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
670	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
671	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0

Correlation Matrix

movieId	1	2	3	4	5	6	7	9
movieId								
1	1.000000	0.223742	0.183266	0.071055	0.105076	0.201503	0.156075	0.023699
2	0.223742	1.000000	0.123790	0.125014	0.193144	0.085889	0.117211	0.053810
3	0.183266	0.123790	1.000000	0.147771	0.317911	0.158071	0.390331	0.274638
4	0.071055	0.125014	0.147771	1.000000	0.150562	0.024466	0.156876	0.238193
5	0.105076	0.193144	0.317911	0.150562	1.000000	0.186936	0.339605	0.339402
...
170	0.153706	0.235887	0.121831	0.071459	0.083620	0.141157	0.086985	-0.033039
172	0.054648	0.262609	-0.018163	0.055141	0.098029	0.063131	-0.023443	0.035510
173	0.058416	0.359524	0.061617	-0.010186	0.059107	0.171023	0.011611	0.092491
175	0.063768	0.068592	-0.001141	0.027405	0.037671	0.120118	-0.018445	0.001659
176	0.179861	0.098706	0.076697	0.119778	0.056150	0.126835	0.132488	0.087999

“

Based on another motive id and its
rating => new recomandation

MOVIE ID : 5
RATING: 1

3	5.000000
802	2.295606
719	2.194623
494	2.020616
7	1.951654

- *Size of trainset: 80003*
- *Size of testset: 20001*

BEFORE OPTOIMIZATION

k=40, min_k=1

RMSE on training data : 0.71

RMSE on test data: 0.96

KNN

USER TO USER

AFTER OPTIMIZATION

k = 15, min_k= 5

RMSE on training data : 0.73

RMSE on test data: 0.96

- *Size of trainset: 80003*
- *Size of testset: 20001*

BEFORE TUNING

**n_factors': 100, 'reg_all': 0.02, '
n_epochs': 20**

RMSE on training data : 0.6431

RMSE on test data: 0.9023

SVD

Single value decomposition

AFTER TUNING

**'n_factors': 100, 'reg_all': 0.06,
'n_epochs': 48**

RMSE on training data : 0.669

RMSE on test data: 0.885

Content Based Recommender

CREDITS

KEYWORDS

Movie Cast, Crew, Keywords

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

improved_recommendations('The DUFF')					
	title	vote_count	vote_average	year	wr
1547	The Breakfast Club	2189	7	1985	6.709602
5207	Mean Girls	2401	6	2004	5.884404
2181	American Pie	2358	6	1999	5.882624
7932	Friends with Benefits	2306	6	2011	5.880396
8906	Paper Towns	1997	6	2015	5.865193
8440	The Spectacular Now	1074	6	2013	5.782682
3712	The Princess Diaries	1063	6	2001	5.781086
7134	High School Musical 3: Senior Year	858	6	2008	5.746351
7494	American Pie Presents: The Book of Love	454	5	2009	5.119690
7755	Beastly	878	5	2011	5.081010

Hybrid Recommender

Input: User ID and the Title of a Movie

Output: Similar movies sorted on the basis of expected ratings by that particular user.

hybrid(23, 'Dr. No')

	title	vote_count	vote_average	year	id	est
673	Mission: Impossible - Rogue Nation	3274.0	7.1	2015	177677	3.802092
383	From Russia with Love	773.0	6.9	1963	657	3.784286
918	The Spy Who Loved Me	515.0	6.6	1977	691	3.723132
638	Casino Royale	3930.0	7.3	2006	36557	3.670023
916	On Her Majesty's Secret Service	464.0	6.5	1969	668	3.597211
226	You Only Live Twice	541.0	6.5	1967	667	3.574294
561	Mission: Impossible	2677.0	6.7	1996	954	3.560654
204	Diamonds Are Forever	562.0	6.3	1971	681	3.556623
785	Wait Until Dark	112.0	7.5	1967	11206	3.552732
230	Never Say Never Again	348.0	5.8	1983	36670	3.547374

hybrid(500, 'True Lies')

	title	vote_count	vote_average	year	id	est
522	Terminator 2: Judgment Day	4274.0	7.7	1991	280	3.525970
7049	Get Smart	1072.0	6.0	2008	11665	3.436580
1011	The Terminator	4208.0	7.4	1984	218	3.433977
561	Mission: Impossible	2677.0	6.7	1996	954	3.182766
3204	Diamonds Are Forever	562.0	6.3	1971	681	3.180613
1304	The Peacemaker	263.0	5.8	1997	6623	3.157220
8205	The Cold Light of Day	277.0	4.8	2012	77948	3.123546
8057	Haywire	508.0	5.6	2011	70435	3.119229
2384	Dr. No	953.0	6.9	1962	646	3.110056
922	The Abyss	822.0	7.1	1989	2756	3.103737



OVERVIEW

Surprise is a Python scikit for building and analyzing recommender systems that deal with explicit rating data.

Thank you for your
attention!

