Movie Recommendation System

Ian McSpiritt
Julianne Lin

Introduction

- Consumer research suggests that a typical streaming platform user will lose interest after 60 to 90 seconds of deciding before selecting a movie title or exiting the platform altogether (Gomez-Uribe & Hunt, 2016)
- Age of influence and curation
- Decision paralysis amongst users
- A great recommender retains users and differentiates from other movie streaming services

Background

Goal

Recommend Movies based on users history and movie attributes

Find Data

Find data that has user history as well as attributes on each movie

Lit Review

Research algorithms that are commonly used in recommender systems

Cleaning

Researching Methods to clean and merge data

Implement Algorithm

Explore classification algorithms and other supervised such as kNN methods

Evaluate

Figure out method to evaluate results using loss functions (Cosine Similarity, RMSE)

Literature Review

Comparative Analysis of Clustering Techniques for Movie Recommendation

They fed datasets into Euclidean distance based algorithms like k-means to cluster them based on specific feature (Aditya et al., 2018)

Feature Selection for Movie Recommendation

They used collaborative recommendation to predict the rating for instances where the user has not seen a certain movie by basing it on the user's neighbors' opinions (ÇATALTEPE et al., 2016)

Data

Movie Lens: Ratings

	userID	movieLensID	rating
0	1	1	4.0
1	1	3	4.0
2	1	6	4.0
3	1	47	5.0
4	1	50	5.0

TMDB

<table>
<tr><th>TMDbID</th><th>IMDbID</th><th>popularity</th><th>budget</th><th>revenue</th><th>title</th><th>cast</th><th>director</th><th>keywords</th></tr>
<tr><td>135397</td><td>tt0369610</td><td>32.985763</td><td>150000000</td><td>1513528810</td><td>Jurassic World</td><td>Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...</td><td>Colin Trevorrow</td><td>monster|dna|tyrannosaurus rex|velociraptor|island</td></tr>
<tr><td>76341</td><td>tt1392190</td><td>28.419936</td><td>150000000</td><td>378436354</td><td>Mad Max: Fury Road</td><td>Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...</td><td>George Miller</td><td>future|chase|post-apocalyptic|dystopia|australia</td></tr>
<tr><td>262500</td><td>tt2908446</td><td>13.112507</td><td>110000000</td><td>295238201</td><td>Insurgent</td><td>Shailene Woodley|Theo James|Kate Winslet|Ansel...</td><td>Robert Schwentke</td><td>based on novel|revolution|dystopia|sequel|dyst...</td></tr>
<tr><td>140607</td><td>tt2488496</td><td>11.173104</td><td>200000000</td><td>2068178225</td><td>Star Wars: The Force Awakens</td><td>Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...</td><td>J.J. Abrams</td><td>android|spaceship|jedi|space opera|3d</td></tr>
<tr><td>168259</td><td>tt2820852</td><td>9.335014</td><td>190000000</td><td>1506249360</td><td>Furious 7</td><td>Vin Diesel|Paul Walker|Jason Statham|Michelle ...</td><td>James Wan</td><td>car race|speed|revenge|suspense|car</td></tr>
</table>

Movie Lens: Links

	movieLensID	IMDbID	TMDbID
0	1	tt0114709	862
1	2	tt0113497	8844
2	3	tt0113228	15602
3	4	tt0114885	31357
4	5	tt0113041	11862

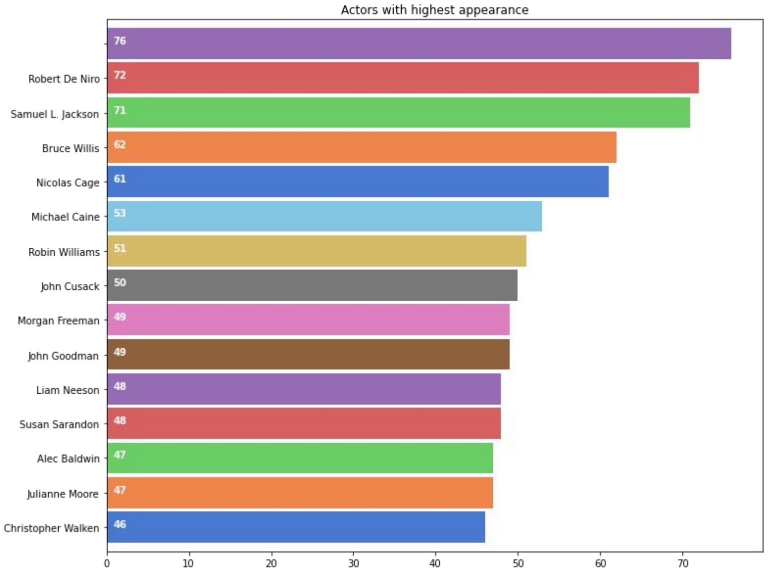
IMDB

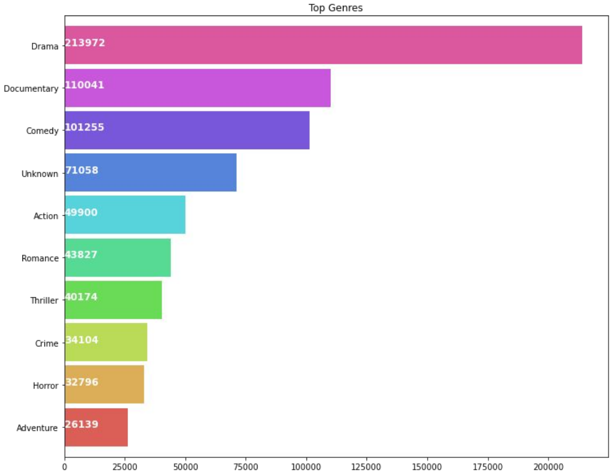
	IMDbID	type	title	originalTitle	adult	year	endYear	runtime	genres
0	tt0000001	short	Carmencita	Carmencita	0	1894	\N	1	Documentary,Short
1	tt0000002	short	Le clown et ses chiens	Le clown et ses chiens	0	1892	\N	5	Animation,Short
2	tt0000003	short	Pauvre Pierrot	Pauvre Pierrot	0	1892	\N	4	Animation,Comedy,Romance
3	tt0000004	short	Un bon bock	Un bon bock	0	1892	\N	12	Animation,Short
4	tt0000005	short	Blacksmith Scene	Blacksmith Scene	0	1893	\N	1	Comedy,Short

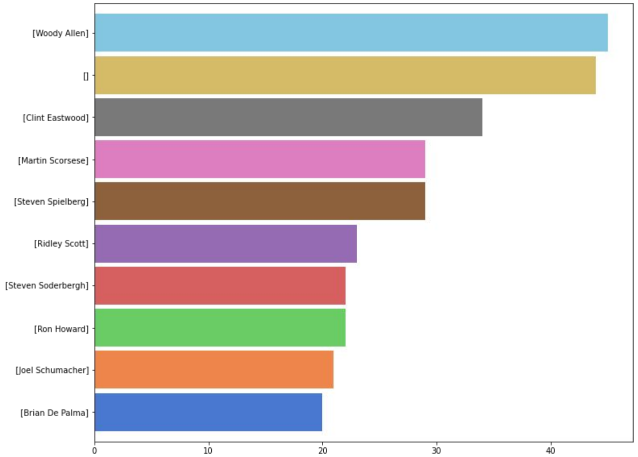
Data

movieLensID	IMDbID	title	genres	cast	director	keywords
1	tt0114709	Toy Story	[0, 0, 1, 1, 0, 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, 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	tt0113497	Jumanji	[0, 0, 1, 0, 0, 1, 0, 0, 0, 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, 0, 0, 0, 0, 0, 0, ...	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
3	tt0113228	Grumpier Old Men	[0, 0, 0, 0, 0, 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, 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	tt0114885	Waiting to Exhale	[0, 0, 0, 0, 0, 1, 0, 0, 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, 0, 0, 0, 0, 0, 0, 0, ...	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
5	tt0113041	Father of the Bride Part II	[0, 0, 0, 0, 0, 1, 0, 0, 0, 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, 0, 0, 0, 0, 0, 0, ...	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...

Data Explored







Approach

Explore each data set and attributes in each

Step 1

Step 2

Clean Data to merge and encode attributes that we intuitively selected

Use KNN by comparing users top rated movies recommending similar movies with strong cosine similarity

Step 3

Step4

Evaluate Cosine similarity and RMSE of top rated movies compared to recommendations

Results

```
Hello! What is your user id? (num between 1 and 610): 246
You've rated 161 movies. How many of your top rated movies would you like us to compare?: 5
How many movies would you like us to recommend?: 10
```

	IMDbID	similarity	RMSE
22432	tt0167261	0.050000	0.003983
22113	tt0120737	0.100000	0.005633
25389	tt1170358	0.447938	0.091168
24884	tt0903624	0.447938	0.091168
27412	tt2310332	0.447938	0.091168
46192	tt2194499	0.468784	0.077678
4751	tt0388795	0.500000	0.014638
23424	tt0360717	0.500000	0.014638
3022	tt0119349	0.523223	0.060121
34514	tt1014759	0.533333	0.079686

	IMDbID	title	genres	cast	director	keywords
3023	tt0119349	The Ice Storm	[Drama]	[Kevin Kline, Joan Allen, Sigourney Weaver, He...	[Ang Lee]	[based on novel, 1970s, thanksgiving, dysfunct...
3183	tt0120737	The Lord of the Rings: The Fellowship of the Ring	[Action, Adventure, Drama]	[Elijah Wood, Ian McKellen, Viggo Mortensen, L...	[Peter Jackson]	[elves, dwarves, orcs, middle-earth (tolkien),...
3503	tt0167261	The Lord of the Rings: The Two Towers	[Action, Adventure, Drama]	[Elijah Wood, Ian McKellen, Viggo Mortensen, L...	[Peter Jackson]	[elves, orcs, middle-earth (tolkien), hobbits,...
4495	tt0360717	King Kong	[Action, Adventure, Drama]	[Naomi Watts, Jack Black, Adrien Brody, Thomas...	[Peter Jackson]	[film business, screenplay, show business, fil...
4752	tt0388795	Brokeback Mountain	[Drama, Romance]	[Heath Ledger, Jake Gyllenhaal, Randy Quaid, M...	[Ang Lee]	[gay, countryside, homophobia, loss of lover, ...
5955	tt0903624	The Hobbit: An Unexpected Journey	[Adventure, Fantasy]	[Ian McKellen, Martin Freeman, Richard Armitag...	[Peter Jackson]	[riddle, elves, dwarves, orcs, middle-earth (t...
6120	tt1014759	Alice in Wonderland	[Adventure, Family, Fantasy]	[Mia Wasikowska, Johnny Depp, Anne Hathaway, H...	[Tim Burton]	[based on novel, fictional place, queen, alice...
6460	tt1170358	The Hobbit: The Desolation of Smaug	[Adventure, Fantasy]	[Martin Freeman, Ian McKellen, Richard Armitag...	[Peter Jackson]	[elves, dwarves, orcs, middle-earth (tolkien),...
8333	tt2194499	About Time	[Comedy, Drama, Fantasy]	[Rachel McAdams, Bill Nighy, Domhnall Gleeson,...	[Richard Curtis]	[london, father-son relationship, time travel]
8483	tt2310332	The Hobbit: The Battle of the Five Armies	[Adventure, Fantasy]	[Martin Freeman, Ian McKellen, Richard Armitag...	[Peter Jackson]	[corruption, elves, dwarves, orcs, middle-eart...

Conclusion

- Our predictions had very 'good' cosine similarities and RMSE and we believe it could potentially be overfitted to the top rated movies
- Potential to cluster user's and recommend based on similar users opposed to focusing more on content of instances
- Expansion of our training set to include not just highly rated movies by the particular users
- Recommendations were strong

Time Log

Julianne			
date	time	hours	task
10/13	10-11	1	Data Search
10/16	9-10	1	Literature Search
10/17	5-7	2	Literature Review
10/20	5:30-6:30	1	Meet with teaching team
10/26	5:30-6:30	1	Project Proposal Planning
10/27	12-2	2	Literature Search
10/29	6-8	2	Literature Review
10/30	1-3	2	Literature Review
10/31	3:30-5:30	2	Literature Review
10/31	5:30-6:30	1	Project Proposal Writing
10/31	9-12	3	Project Proposal & Presentation Prep
11/2	5-6	1	Project Proposal Final Details & Presentation Prep
11/9	5-6	1	Project Timeline Planning
11/14	12-2	2	Data Accessing
11/16	5-6	1	Data Accessing & Meeting
11/30	7-9	2	NLP Research & Meeting
12/7	7-10	3	Data Accessing & Cleaning & Meeting
12/8	5-7	2	Data Cleaning
12/9	7-10	3	Data Cleaning
12/10	1-4	3	Data Cleaning & Meeting
12/11	10-12	2	Data Cleaning & Algorithm Implementation
12/12	2-4	2	Data Cleaning & Algorithm Implementation
12/12	6-7	1	Data Cleaning & Algorithm Implementation
12/12	10-11	1	Algorithm Implementation & Meeting
12/13	12-2	2	Data Cleaning & Transformation
12/13	4-6	2	Data Cleaning & Algorithm Implementation
12/14	2-4	2	Algorthim Implementation
12/14	10-12	2	Algorithm Implementation & Meeting
12/15	2-3	1	Data Visualization

Ian			
date	time	hours	task
10/13	10-11	1	Data Search/Meeting
10/17	8-10	2	Api Research on TMDB
10/20	4:30-6:30	1	Meeting and Data exploration
10/29	12-2	2	Working on Proposal
10/30	12-4	4	Working on Proposal
10/31	12-5	5	Working on Proposal
10/31	9-11	2	Working on Proposal
11/22	5-9	4	Data Cleaning and Meeting
11/29	5-9	4	Data Cleaning and Meeting
12/5	11-5	6	Data Cleaning
12/7	7-10	3	Data Accessing & Cleaning & Meeting
12/10	1-4	3	Data Cleaning & Meeting
12/11	12-5	5	Data Cleaning & Algorithm Implementation
12/12	6-11	5	Data Cleaning & Algorithm Implementation & Meeting
12/13	10-1	3	Algorithm Implementation/ Feature testing
12/14	8-12	4	Evaluation and presentation
12/15	3:30-6:30	3	Presentation and Prep

References

Aditya, T. S., Rajaraman, K., & Monica Subashini, M. (2018). Comparative analysis of Clustering Techniques for movie recommendation. MATEC Web of Conferences, 225, 02004. https://doi.org/10.1051/matecconf/201822502004

Gomez-Uribe, C. A., & Hunt, N. (2016). The Netflix Recommender System. ACM Transactions on Management Information Systems, 6(4), 1–19. https://doi.org/10.1145/2843948

ÇATALTEPE, Z., ULUYAĞMUR, M., & TAYFUR, E. (2016). Feature selection for movie recommendation. TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES, 24, 833–848. https://doi.org/10.3906/elk-1303-189

Thanks!

Do you have any questions?

ian.mcspiritt@rutgers.edu
jjl302@scarletmail.rutgers.edu