Anime Recommender System

Abishek Ganesh

Aditya Trilok Muralidharan

Padmavathi Karunaiananda Sekar

Under the guidance of Dr. Sridhar Nerur

Problem Description

- Focus on building the best recommendation system for anime fans
- Using Alternating Least Squares & Similarity within genre for recommendation
- For each user ID, we recommend the top 10 animes based on user preference

Dataset Description

- Dataset is from Kaggle and the competition is active
- Dataset contains information on 73,516 users on 12,294 animes
- We have around 7 million records for user data and around 12000 records for anime
- The recommendation is based on “rating” for ALS & “genre” for Cosine Similarity
- Preprocessing was performed on the original dataset to apply the algorithms

Dataset Description - anime.csv

- anime_id - myanimelist.net's unique id identifying an anime
- name - full name of anime
- genre - comma separated list of genres for this anime
- type - movie, TV, OVA, etc
- episodes - how many episodes in this show (1 if movie)
- rating - average rating out of 10 for this anime
- members - number of community members that are in this anime's "group"

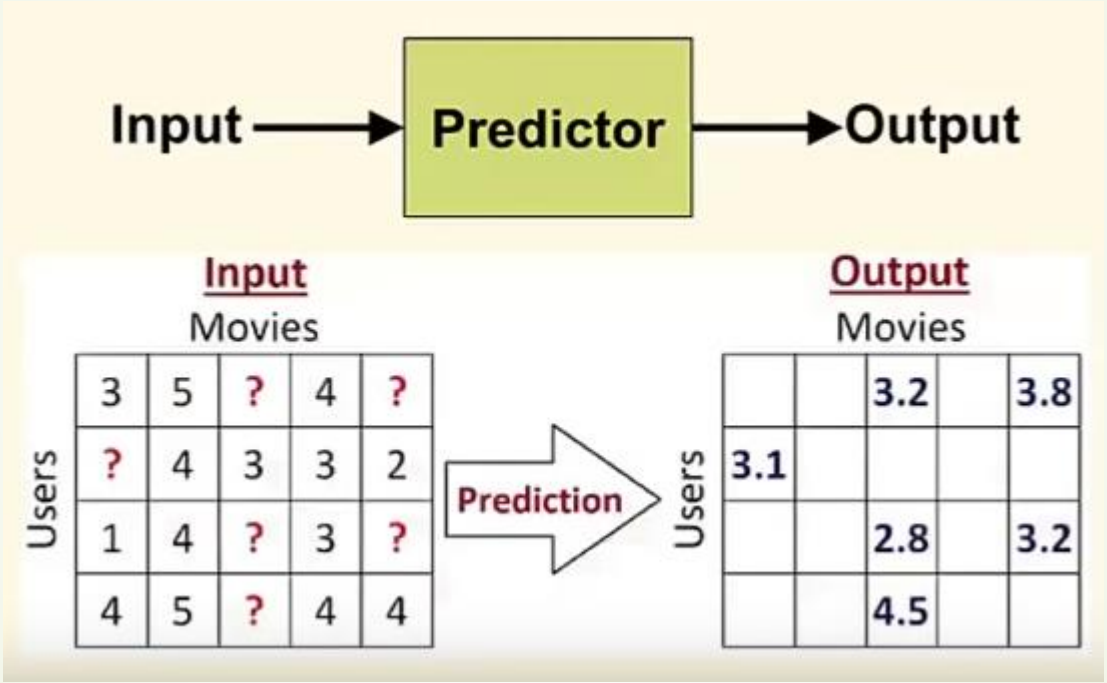
Dataset Description - rating.csv

- user_id - non identifiable randomly generated user id
- anime_id - the anime that this user has rated
- rating - rating out of 10 this user has assigned (-1 if the user watched it but didn't assign a rating)

Data Preprocessing

- Removed "name" column & deleted missing values from "episodes", "genre" and "ratings" column in anime.csv
- Removed rating = -1 from rating.csv
- anime_name.csv contains anime_id and name for mapping purpose

Alternating Least Squares



ALS Metrics

alpha	nonnegative	rank	maxIter	regParam	RMSE
1	FALSE	10	40	0.1	1.125891
1	FALSE	10	50	0.1	1.124501
1	FALSE	10	60	0.1	1.12264
1	FALSE	10	75	0.1	1.123311
1	FALSE	10	90	0.1	1.124913
1	FALSE	10	100	0.1	1.123787
1	FALSE	30	50	0.1	1.1112
1	FALSE	50	50	0.1	1.107503
1	FALSE	100	10	0.1	1.151631
1	TRUE	20	40	0.1	1.115722
1	TRUE	20	40	0.1	1.1188285
1	TRUE	15	40	0.1	1.122352

ALS Metrics – with Cross Validation

maxIter	regParam	RMSE
[10,5,2]	[0.1,0.2,0.3]	1.154694
[2,3,4]	[0.1,0.2,0.3]	1.207455

Recommendation for user_id 11024

user_id	anime_id	rating	prediction	anime_id	name
14454	1639	9	14.574396	1639	Boku no Pico
42309	1535	8	11.523261	1535	Death Note
32172	28977	10	11.360651	28977	Gintama°
66941	820	9	11.319762	820	Ginga Eiyuu Densetsu
59134	1575	10	11.262725	1575	Code Geass: Hangy...
63126	11757	10	11.223665	11757	Sword Art Online
13370	1	10	11.17156	1	Cowboy Bebop
37358	19815	10	11.09791	19815	No Game No Life
15541	9969	10	11.07065	9969	Gintama'
59579	19	10	11.070261	19	Monster
46035	11757	10	11.069681	11757	Sword Art Online

Cosine Similarity

$$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

Cosine Similarity Metrics

	0	1	2	3
0	987	1.000000	987	Dragon Ball GT: Goku Gaiden! Yuuki no Akashi w...
1	904	1.000000	904	Dragon Ball Z Movie 11: Super Senshi Gekiha!! ...
2	22777	1.000000	22777	Dragon Ball Kai (2014)
3	25389	1.000000	25389	Dragon Ball Z Movie 15: Fukkatsu no F
4	22695	1.000000	22695	Dragon Ball Z: Summer Vacation Special
5	6033	1.000000	6033	Dragon Ball Kai
6	6714	1.000000	6714	Dragon Ball Z: Atsumare! Gokuu World
7	813	1.000000	813	Dragon Ball Z
8	903	0.935414	903	Dragon Ball Z Movie 10: Kiken na Futari! Super...
9	22699	0.935414	22699	Dragon Ball Z: Zenbu Misemasu Toshi Wasure Dra...
10	12231	0.935414	12231	Dragon Ball: Episode of Bardock
11	223	0.925820	223	Dragon Ball
12	3848	0.925820	3848	One Piece Movie 9: Episode of Chopper Plus - F...
13	8740	0.925820	8740	One Piece Film: Strong World Episode 0
14	14837	0.925820	14837	Dragon Ball Z Movie 14: Kami to Kami
15	1094	0.925820	1094	One Piece: Umi no Heso no Daibouken-hen
16	460	0.925820	460	One Piece Movie 2: Nejimaki-jima no Daibouken
17	16239	0.925820	16239	One Piece: Episode of Luffy - Hand Island no B...
18	32051	0.925820	32051	One Piece: Adventure of Nebulandia
19	1237	0.925820	1237	One Piece: Oounabara ni Hirake! Dekkai Dekkai ...

Other Similarity Metrics

- Jaccard Similarity
- Euclidean Distance
- Manhattan Distance

Inferences

anime_id1	anime_id2	cosine_similarity	jaccard_similarity	manhattan_distances	euclidean_distances
32281	32281	1.0	1.0	0.0	0.0
32281	5114	0.1889822365046136	0.7906976744186046	27167.0	27167.0
32281	28977	0.0	0.7441860465116279	3304.0	3304.0
32281	9253	0.0	0.8604651162790697	23028.0	23028.0
32281	9969	0.0	0.7441860465116279	22312.0	22312.0
32281	32935	0.4472135954999579	0.8837209302325582	654.0	654.0
32281	11061	0.0	0.813953488372093	21220.0	21220.0
32281	820	0.25	0.8604651162790697	31461.0	31461.0
32281	15335	0.0	0.7441860465116279	16946.0	16946.0
32281	15417	0.0	0.7441860465116279	16864.0	16864.0
32281	4181	0.6708203932499369	0.9302325581395349	28100.0	28100.0
32281	28851	0.5773502691896258	0.9302325581395349	3430.0	3430.0
32281	918	0.0	0.7441860465116279	31363.0	31363.0
32281	2904	0.20412414523193154	0.813953488372093	29377.0	29377.0
32281	28891	0.4472135954999579	0.8837209302325582	3390.0	3390.0
32281	199	0.5773502691896258	0.9302325581395349	32082.0	32082.0
32281	23273	0.6708203932499369	0.9302325581395349	9008.0	9008.0
32281	24701	0.1889822365046136	0.7906976744186046	7580.0	7580.0

Challenges

- Number of records in consideration for ALS & Item based similarity
- N * N item matrix for similarity is too huge (ex: 11830^2 ~= 140 Million)
- Regularization parameter tuning

Future Scope

- Handling data with respect to viewed but unrated animes
- Inclusion of other predictors to generate recommendations
- The similarity based recommendation can be extended to other textual similarities between 2 items (2 animes)
- Data for the plot can be extracted from forum discussions

References

- https://spark.apache.org/docs/preview/ml-collaborative-filtering.html
- https://www.elenacuoco.com/2016/12/22/alternating-least-squares-als-spark-ml/
- http://alex.smola.org/teaching/10-701-15/recommender.html
- https://www.youtube.com/watch?v=TSv6eLAOt78
- https://www.kaggle.com/CooperUnion/anime-recommendations-database
- http://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine_similarity.html
- http://scikit-learn.org/stable/modules/metrics.html
- http://scikit-learn.org/stable/modules/generated/sklearn.metrics.jaccard_similarity_score.html