Lan Mei
BAX 452 HW Assignment #7
HW#7 Anime Recommendation.ipynb

Overview

Anime Recommendation is a project that create collaborative filtering recommendation using SVD in Surprise framework. The recommendation is based on user preference data from 25,451 users on 12,294 animes.
Data Source: Anime Recommendation - Kaggle

Steps of Analysis

- **Data Preprocessing**
  Original dataset contains Anime.csv and Rating.csv, which is about the information of animes and user rating. After data cleaning, we got table *rating* and *anime_mapping*:

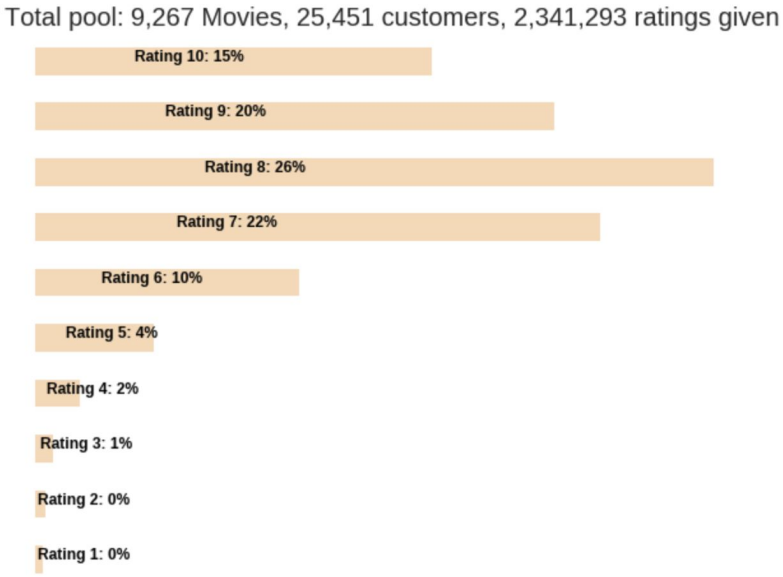
	user_id	anime_id	rating
0	1	8074	10
1	1	11617	10
2	1	11757	10

Rating

	anime_id	name	genre	type
0	32281	Kimi no Na wa.	Drama, Romance, School, Supernatural	Movie
1	5114	Fullmetal Alchemist: Brotherhood	Action, Adventure, Drama, Fantasy, Magic, Mili...	TV
2	28977	Gintama°	Action, Comedy, Historical, Parody, Samurai, S...	TV

Anime_mapping

From the barplot of the dataset on the right, we can find that the rating is lower-rating-skewed. It seems that users are generous about their rating, with "8" as the most frequent rating score.



- **Collaborative Filtering Recommendation Exploration**
  I used singular value decomposition (SVD) — one of the Matrix Factorization models for identifying latent factors. With 3-fold cross-validation, we got the mean RMSE/MAE.

```
------------------------------------------------------------------------
Fold 1          Fold 2          Fold 3          Mean RMSE: 1.1749
RMSE: 1.1752    RMSE: 1.1741    RMSE: 1.1752    Mean MAE : 0.8824
MAE:  0.8824    MAE:  0.8819    MAE:  0.8831
```

  If we take user_id = 45 as an example:

user_id	rating	name	genre	type
45	10	Ouran Koukou Host Club	Comedy, Harem	TV
45	9	Ghost Hunt	Comedy, Horror	TV
45	9	Vampire Knight	Drama, Mystery	TV
45	9	Skip Beat!	Comedy, Drama	TV
45	10	Angel Beats!	Action, Comedy	TV
45	10	Kaichou wa Maid-sama!	Comedy, Romance	TV
45	9	Chihayafuru	Drama, Game	TV
45	9	Another	Horror, Mystery	TV
45	9	Sword Art Online	Action, Adventure	TV

Similar genre

What user_id = 45 offered high ratings

Two overlapped animes

Various type

name	genre	type	Estimate_Score
Kimi no Na wa.	Drama, Romance	Movie	9.564
Ouran Koukou Host Club	Comedy, Harem	TV	9.469
Gintama Movie: Kanketsu-hen	Action, Comedy	Movie	9.462
Howl no Ugoku Shiro	Adventure, Drama	Movie	9.452
Angel Beats!	Action, Comedy	TV	9.418
There She Is!!	Comedy, Romance	ONA	9.373
Sen to Chihiro no Kamikakushi	Adventure, Drama	Movie	9.368
xxxHOLiC	Comedy, Drama	TV	9.292
Ginga Eiyuu Densetsu	Drama, Military	OVA	9.290
Gintama': Enchousen	Action, Comedy	TV	9.285

Recommendation

Conclusion Based on the Evaluation and Example

- Collaborative Filtering collects and analyzes past user behavior, and predicts what users like based on their similarity to other users. The accuracy can be improved with larger dataset on users/ items.
- From the example we can find that the recommendation offers various choices with only two anime that the user has already watched. Though the recommended animes have types that are largely different from the user's preference(mainly "TV"), they are still from the genres that the user is interested in.
- With Collaborative Filtering we may discover the possible interests of users.