Anime Recommendation System Report

MAS 651

By

Kaiting Liang, Qihan Liu, Yina Luo

Table of Content

1 Project Overview

Our project aims to develop an anime recommendation system using a dataset obtained from Kaggle, and based on user rating on My Anime List website (an anime rating website). This project is especially appealing to us because we are strongly interested in watching anime. We are excited to create a system that provides users with personalized anime recommendations based on their preferences. To achieve this, we will focus on using collaborative filtering methods. In our evaluation process, we evaluated seven different models provided by the surprise package, consisting of two user-based KNN models, two item-based KNN models, to ensure that we select the most effective model for our recommendation system.

2 Data Summary

2.1 Data Description

The dataset we are working with collects information on user preference data from 73,516 users on 12,294 anime. There are two csv files, ‘anime’ and ‘rating’. The anime file contains 12,294 records under 7 attributes, including the anime_id, name, genre, type, episodes, rating and members. The rating file contains 7,813,737 records with 3 attributes, including user_id, anime_id, and rating. By merging these two datasets, we have created a comprehensive collection of user preference data, which allows us to better train the model.

2.1.1 Anime data set

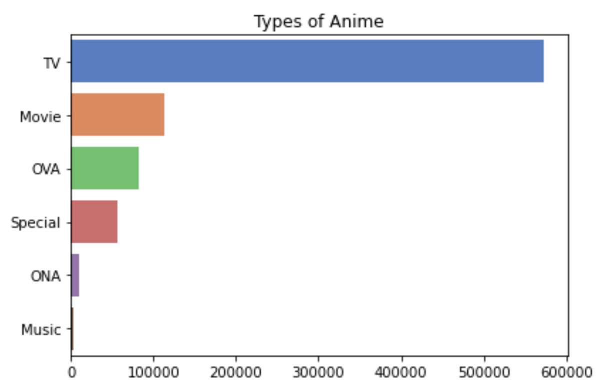
Attributes	**Description**
anime_id	Unique anime identifier
name	Full name of the anime
genre	Animation genre(comedy)
type	Animation type(movie, TV, OVA)
episodes	Number of episodes(1 indicate movie)
rating	Average rating of anime(out of 10)
members	Number of community members in this anime’s group

2.1.2 Rating data set

Attributes	**Description**
user_id	Non-identifiable randomly generated user id
anime_id	Anime rated by the user
rating	User’s rating of the anime out of 10(-1 indicates that the user watched it but did not rate it)

2.2 Data Analysis

2.2.1 Analysis Based on Anime Type and Community Size



	name	members
0	Death Note	1013917
1	Shingeki no Kyojin	896229
2	Sword Art Online	893100
3	Fullmetal Alchemist: Brotherhood	793665
4	Angel Beats!	717796

We found that the majority of the animes in My Anime List were aired on television, followed by anime movies. We also discovered that 'Death Note' has the largest community size on the platform with over 1 million members, indicating a significant level of popularity and engagement among fans.

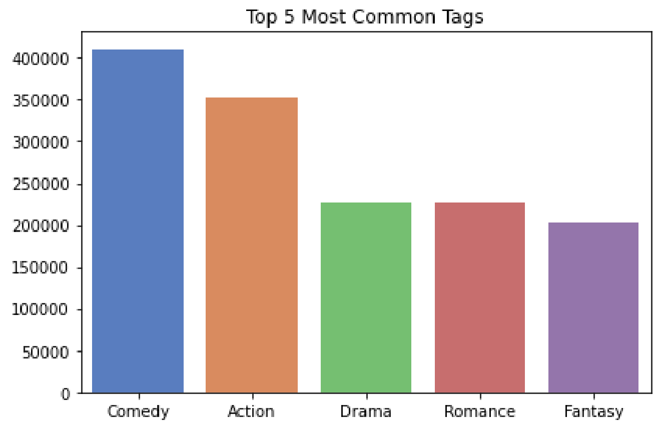
2.2.2 Analysis Based on Anime Rating

	name	rating
0	Kimi no Na wa.	9.37
1	Fullmetal Alchemist: Brotherhood	9.26
2	Gintama°	9.25
3	Steins;Gate	9.17
4	Gintama'	9.16
5	Haikyuu!!: Karasuno Koukou VS Shiratorizawa Ga...	9.15
6	Hunter x Hunter (2011)	9.13
7	Gintama': Enchousen	9.11
8	Ginga Eiyuu Densetsu	9.11
9	Gintama Movie: Kanketsu-hen - Yorozuya yo Eien...	9.10

	name	rating
0	Tenkuu Danzai Skelter+Heaven	2.00
1	Utsu Musume Sayuri	2.14
2	Hametsu no Mars	2.37
3	Nami	2.67
4	Ningen Doubutsuen	2.78
5	Shitcom	2.93
6	Abunai Sisters: Koko & Mika	2.95
7	Tsui no Sora	2.98
8	Bulsajo Robot Phoenix King	3.11
9	Aki no Puzzle	3.27

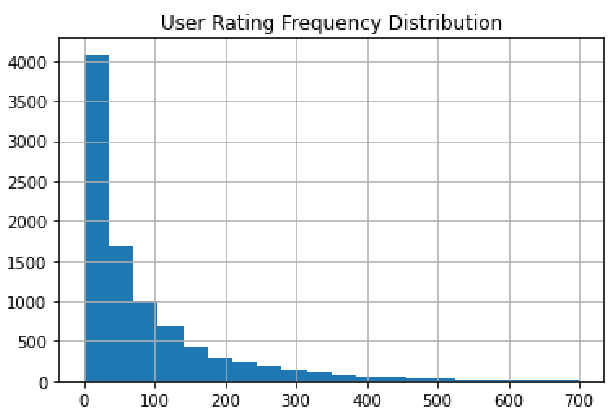
When examining the anime ratings, we discovered that the top 10 ranked animes all had ratings above 9, while the worst 10 animes were ranked below 3.5.

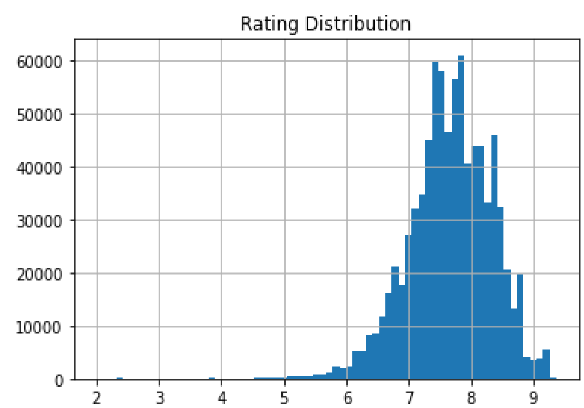
2.2.3 Analysis Based on Anime Genre



We also analyzed the genres of each anime to gain insight into the audience's preferences, and found that the Comedy genre is the most popular, followed by action and drama genres. This was further supported by a Word Cloud and bar chart analysis.

2.2.4 Analysis Based on User Behaviors





By examining the rating behavior of users, we found that the distribution was right-skewed, with most users rating only one anime and very few rating more than 500 anime. The distribution of the ratings revealed that the most common score awarded to an anime falls within the range of 7 to 8.

3 Model Building

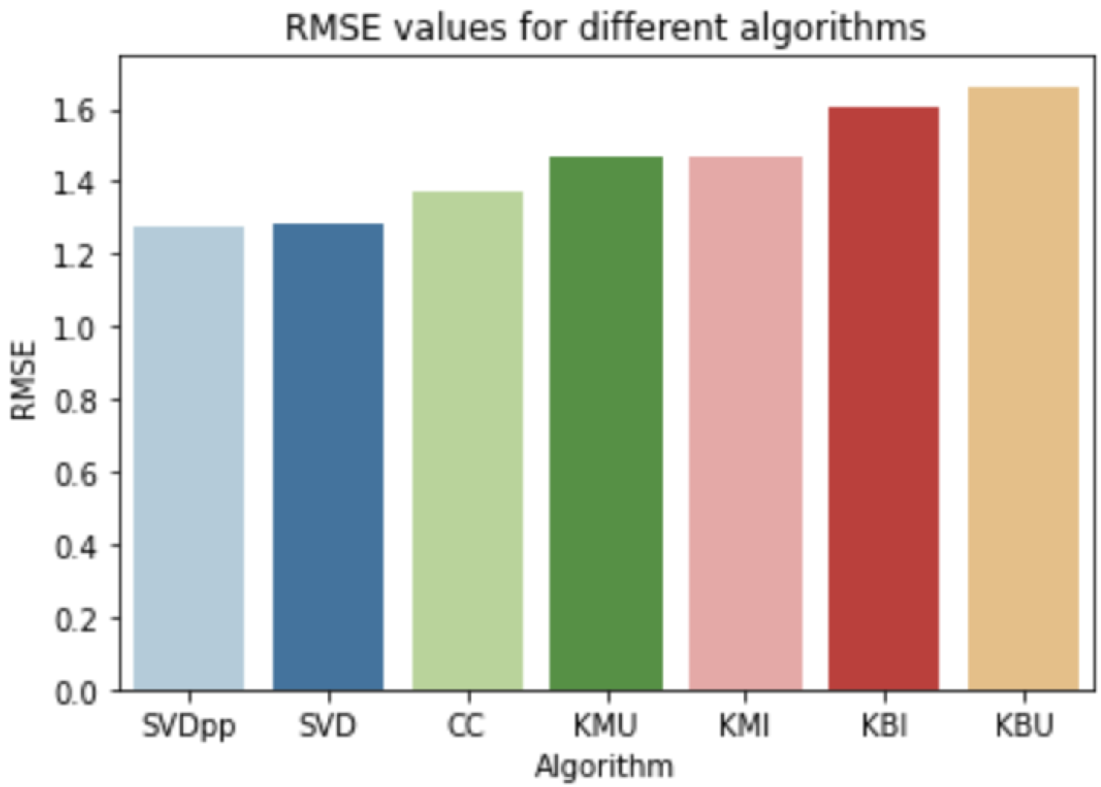
3.1 Model Overview

Our anime recommendation system primarily employs collaborative filtering, where we evaluated seven different models using the surprise package. These models include two user-based KNN models, two item-based KNN models, SVD, SVDpp, and Co-Clustering models. To explore new models, we use SVDpp and Co-Clustering, which have not been used before for us. SVD, is a matrix factorization technique that decomposes the user-item rating matrix into a lower-dimensional representation. SVDpp, is an extension of SVD, considers implicit user feedback and captures more information about user preferences, making it a suitable fit for our system. CoClustering is a clustering algorithm that groups users and items based on their similarity in the rating matrix. We cleaned and filtered our dataset, reducing it to 800,000 entries to fit our models.

3.2 Model Performance and Selection

Model	RMSE
SVDpp	1.2784
SVD	1.2857
CoClustering	1.3676
User-based KNN Means	1.466291
Item-based KNN Means	1.466291
Item-based KNN Basic	1.602499
User-based KNN Basic	1.663005

As shown in the table provided, the SVDpp model had the lowest RMSE and was the top-performing model. This is an indication that the SVDpp is a better choice for the application.



We fine-tuned its hyperparameters, get the best parameter with {'n_factors': 20, 'n_epochs': 200, 'lr_all': 0.002, 'reg_all': 0.4}, resulting in better results of **RMSE 1.25** during testing.

3.3 Model Testing

Our final model accurately identified users' preferences and recommended content similar to their viewing history. During testing, we observed that the top 10 recommended animes for a user were often animes they had watched or were currently watching, demonstrating the model's ability to accurately identify users' preferences and recommend relevant content.

4 Business insight

Recommendation systems have become increasingly popular in commercial applications, particularly in the entertainment industry. Our anime recommendation system aims to provide high-quality reviews to viewers, helping them determine whether an anime is worth watching or not. Personalized recommendations based on viewing history and rating can significantly improve user experience and increase customer satisfaction and retention. Our system has the potential to provide valuable insights into users' behavior and preferences, such as genre preference and rating distribution, which can be used for targeted marketing and advertising campaigns.

4.1 Business application

For My Anime List website, a recommendation system can help to analyze the market through users' preferences and offer similar products to users. This can help My Anime List website better understand their audience and tailor their products to meet their needs, resulting in increased customer satisfaction and loyalty. Our anime recommendation system has the potential to benefit both users and businesses alike by providing personalized recommendations and valuable insights into user behavior and preferences.

4.2 Further Work

An additional direction for future research is to integrate contextual information into the recommendation system. One approach to achieve this could be by employing content-based filtering and assessing its performance. Alternatively, a hybrid model that combines collaborative filtering with content-based filtering could be explored to enhance the recommendation accuracy.