Anime Recommendation

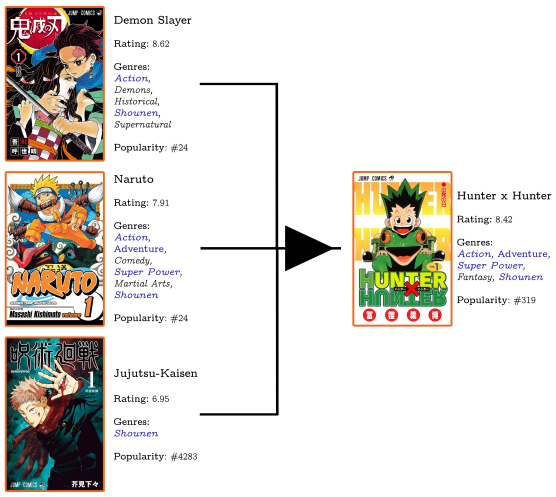
Ryan Knobel * Luis Ruiz † Mike Panuelos ‡ Juan Perez §

Abstract

We were given two data sets containing information pertaining to all anime and a user's past watch history. Using this information we were tasked to design a system that takes this data and provides recommendations to users based on what they enjoy.

1 Introduction

Data mining has contributed to an abundance of valuable information by revealing patterns within large sets of data through the process of finding anomalies, patterns, and correlations, allowing the prediction of outcomes. Many models have been created for streaming services that recommend shows and movies based on a user's history to understand why they watched or didn't watch and enjoyed or didn't enjoy what they viewed. Similarly, if you were given a user's data that shows what anime they watched and how they rated it and also have access to the anime's data such as its overall average rating by all the users that have watched it, the genres it pertains to, its popularity, and etc. Our goal is to develop a model that can determine whether a specific anime is a good recommendation for a user and then use this model to generate a recommendation for the user.



*Department of Computer Science, University of Texas Rio Grande Valley, ryan.knobel01@utrgv.edu

†Department of Computer Science, University of Texas Rio Grande Valley, luis.ruiz02@utrgv.edu

‡Department of Computer Science, University of Texas Rio Grande Valley, mike.panuelos01@utrgv.edu

§Department of Computer Science, University of Texas Rio Grande Valley, juan.m.perez02@utrgv.edu

2 Model Construction

Our initial discussion sought to conclude what data from an anime could best provide insight into the user's experience and explain why the user chose to watch it for the length that they did and rate it the score that they gave. With many contenders such as the anime's rank, the number of episodes, the producer, and the time period that it was aired, we concluded that our initial model should be based on the anime's genre and score (the average rating given by all users). Our model considers two features. The first feature is composed of the user's top 25% favorite genres and the average rating of all the anime that they watched. The second feature is composed of an anime's genres and its score.

3 Testing Procedure

To test the performance of our model, we first construct a testing data set with the provided user history. Each instance of the training data for the model intuitively can be broken into two components: what the user enjoys on average, and information about a specific anime.

3.1 Computing a User's "Average Anime"

Given a user's watch history, we consider every anime that the user rated a 7 or better. From this set, count the appearance of each genre, assigning a value of 1 for the most common genres (top 25 percent) or 0 for all other genres. Store these values in an array, where each position in the array corresponds to a different genre. Additionally, compute the average cumulative rating (across all users) for each anime that the user rated highly and store this value in the array.

UserID	AnimeID	Score	WatchedEpisodes
0	20	8	150
0	38000	9	24
0	38777	10	3

3.2 Computing Specific Anime Information

For any instance of anime, store the genres as binary values in an array, where each position in the array corresponds to different genre and the ordering matches the array described in section 3.1. Similarly, store the cumulative rating for the anime given by all users in this array as well.

3.3 Training Data for the Model

Using the methods described in sections 3.1 and 3.2, it is now possible to construct a feature array that our model can use as input to output whether the anime would be a

good recommendation for the user. Since "good" recommendations are subjective, we define the following procedure to train the model:

- For each individual user:
    1. Compute the user's "Average Anime"
    2. For any anime that the user has rated and watched, compute the respective information
    3. If the user rated the anime above a 7, consider this a good recommendation (1). Otherwise, consider this a bad recommendation (-1)

Given this procedure, we can then create 2 sets of data. The first data set contains the information about the user and each anime, and the second data set contains the corresponding label for the recommendation. The model then intuitively can make predictions for some instance of anime conditionally on what the user enjoys in order to make the best prediction possible.

4 Performance

To test the performance of our proposed model, we split the data set into two different types of sets: a training set (80 percent of the data) and validation set (20 percent of the data). The training set is used to train the model, which includes both the user/anime data and the recommendation label. The validation set on the other hand is used to measure the performance of the model. Here, we provide the model solely with the user/anime data, make predictions using the model, and compare the results to whether the user enjoyed the anime. As a result, performance is decided by the accuracy of the predictions made by the model. Using 300,000 samples total, our model was able to predict whether an anime was a good recommendation for a user with 72 percent accuracy.

5 Conclusion

Our model considers 2 factors when determining whether an anime is a good recommendation for a user. However, there are many other factors that can lead to a user either enjoying or not enjoying the anime. The process of constructing the model revealed this notion, which we leave for future work to potentially yield a higher prediction accuracy. The foundation of the model was further adapted to not only make a prediction for specific anime instances but also provide an anime recommendation that our model determines to be good. As shown in the image in the Introduction and the table in Section 3.1, our model was able to recommend the anime, "Hunter x Hunter" to a user who has watched "Demon Slayer" and scored it an 8, "Naruto" and scored it a 9, and "Jujutsu-Kaisen" and scored it a 10.

References

[1] G. Akutami. Wiki Targeted (Entertainment).

[2] Shueisha. Hunter A Hunter (Volume).

[3] Viz. Naruto (Volume).

□