

Improving MyAnimeList's Recommendations System

Foundations of Artificial Intelligence Final Project Proposal

Fay Feghali

Problem Description

[MyAnimeList](#) (MAL) is the largest online forum for English-speaking anime enthusiasts to rank and discuss their favorite shows. With 150 million page views and over 12 million unique visitors [each month](#), MAL consistently gathers a plethora of data. Since users can request to add shows and movies that are not currently on the website, there are likely very few shows that do not exist as a data point on the site. Even with this enormous user-sourced data set, the recommendations section of MAL leaves much to be desired. Like most everything on the site, recommendations are [user generated](#). Users can choose to write recommendations where they can input two different shows and write why they think that if someone watches one of the shows then they will likely enjoy the second show. MAL lacks a recommendation engine that utilizes its ever growing user base who contribute new information about various anime daily. This project aims to address this issue using various collaborative filtering algorithms to create a user-based recommendation engine.

The main dataset we will be using includes the ratings of anime from the most active MAL users. We will obtain this dataset from scrapping from the MAL website using [MyAnimeList Web Scraper](#). If time permits, a second dataset from the recommendation section of MAL will be used. This data would add to the depth of our analysis because it contains users' personal opinions on whether two anime shows are similar enough to each other to recommend to other users. The user generated recommendation section of MAL is an underutilized dataset in other MAL data analyses and collaborative filtering projects. Various Python libraries like [sklearn](#), [PySpark](#) and [SciPy](#) will be used for the collaborative filtering algorithms.

From a computational perspective, this project starts with running various filtering algorithms on a MAL dataset and then uses the resulting analysis to predict/recommend what anime tv shows individual MAL users would be most interested in. Although some collaborative filtering algorithms include analyzing implicit data, the MAL API does not allow for accessing any implicit user data. Therefore all the algorithms used in this project will be trained on explicit user data like what users input for ratings for each anime and what users input for recommendations. As well as scrapping the data and training our model on the dataset, we will offer a basic interface for users to input their username which will then output a ranked list of recommended anime based on the user's data.

Approach/Algorithms

After extensive research on recommendation engines from Netflix to Spotify, we felt the best course of action would be testing out common collaborative filtering algorithms. We also looked for examples of recommendation algorithms from other anime websites like [Crunchyroll](#), a streaming service dedicated to anime. Using these examples for guidance as well as delving into comparative research studies on various recommendation algorithms, we decided on using various matrix factorization algorithms to create a better recommendation system for MAL.

The algorithms we plan to use to recommend anime to MAL users are the following:

1. [k-Nearest Neighbors](#) (kNN)
2. [Singular Value Decomposition](#) (SVD)
3. [Alternating Least Squares](#) (ALS)
4. [Stochastic Gradient Descent](#) (SGD)

Each of the links above contains an article showing how the algorithm is used for recommendation engines similar to what our goal is for the MAL recommendation engine.

Planned Comparisons

The plan is to compare the four algorithms to each other using regression analyses like Root Mean Square Error (RSME). These methods for comparison will be similar to the methods used in [this study](#). We want to determine which algorithm is the best for recommending anime based on the explicit user data on MAL. Time permitting and if possible, we would also like to create a hybrid recommender system using the best of the various algorithms.

The following studies were also referenced to create this project proposal:

1. [Incorporating Singular Value Decomposition in User-based Collaborative Filtering Technique for a Movie Recommendation System](#)
2. [Collaborative Filtering for Implicit Feedback Datasets](#)
3. [The Netflix Recommender System: Algorithms, Business Value, and Innovation](#)
4. [Matrix Factorization Techniques for Recommender Systems](#)