**INFO7390 Report: Anime Recommender System**

**Zhao Zhang, Liu Bo**

**Source**

https://www.kaggle.com/hernan4444/anime-recommendation-database-2020?select=anime\_with\_synopsis.csv

**Background**

With the rise of eWorld , e-commerce and online advertisement are widely used. Recommender system become unavoidable in our daily life. Many famous webs such as Amazon, Netflix, Youtube all have their own recommender systems. Accurately recommender the items the users want can generate a huge amount of profit for those companies. Nowadays there are different kinds recommender systems, the most famous ones are collaborative filtering and content based filtering. Collaborative filtering find similar items based on user’ past behavior, and recommender possible items that the user may like based on history. Content based recommender is more NLP based and focus on user’ profile or item’ description and recommender similar items. The are also other ways of recommending and many industries are using hybrid recommending which combing multiple ways of recommender systems.

**Objective**

This project aims to have experiments with different recommender systems. The datasets used are from Kaggle, they are anime and rating information on MyAnimeList until the year of 2020. We were trying to implement different kinds of anime recommender.

The focus of this project is to implement collaborative filtering algorithms with surprise library with different algorithms

Except the main task which focus on collaborative filtering, we will also have a taste of other algorithms. Those tasks are optional tasks for us, and they are implemented for the purpose of further exploring and personal interest. Finally, we have done some extra work and use clustering algorithms on anime data to see that anime are clustered.

**Datasets**

There are total 5 datasets on Kaggle and for this project we use only 3 of them. The anime file contains all information about an anime except the synopsis. There are total 34 columns and 48.5k rows. This file is used to read anime’s basic information when output such as anime’s name. To the contrast, anime\_with\_synopsis file contains only basic anime information and its synopsis, there are 5 columns and 48.5 rows. This file is only used in content-based filtering and no other algorithms.

The most import file is rating\_complete.csv, it contains all users’ rating information, and those rating is only valid and included in this file when the user has watched an anime completely (not want to watch or in watching, but watched). There are 3 columns, user\_id, anime\_id, rating, and there are 57 million ratings in total.

**Methodology**

The main task is the mandatory part and we strictly following the project requirements. We tried 6 different collaborative recommender systems with surprise library. Those 6 algorithms are KNN Basic, KNN with Z score, SVD, SVD++, NMF,Coclustering. Each of them use a different statistical and mathematical method to estimate the scores and get the recommended anime. We also tune hyper parameters by grid cross validation and choose the best parameter for each algorithm and compare the results of 6 different algorithms by visualization

Sub tasks are optional part, mainly for personal interest or further exploring. They include other recommend algorithms and also clustering. The recommender algorithms includes most simple recommender: weighted rating, which ranks all anime by not just the averaging the scores, but also weighted each anime based on number of votes for that anime. Second algorithm in this part is contend based filtering, which build tf-idf matrix based on anime’s genres and synopsis and calculate cosine similarities between all anime. Recommendation is item based and will recommend similar anime based one anime’ cosine similarities scores. We will tried neural network to full connect user and anime and calculate scores then recommend.

Finally there is some extra work about clustering, there are 4 different clustering algorithms and they are apply on the anime dataset. The features are extracted only ordinal categories or continuous numbers are kept for the clustering. After conversion, they are all numeric in order to do clustering. The 4 algorithms are K-means, Hierarchical Agglomerative Clustering, DBSCAN, Mean Shift. Visualizations are done for each of them.

**Evaluation**

**Results and Finding**

**Code and Running Instruction**

Tools:

Running environment: Kaggle notebook is the default environment used for this project as our dataset were found on Kaggle and Kaggle provides all the tools and environment needed for this project.

Python Libraries: sklearn, keras, pandas, matplotlib, numpy, seaborn, etc.

Anime\_recommender1: Featured two simple recommender systems: weighted rating and content-based

Anime\_recommender2: Featured 6 recommender systems algorithms: KNN (collaborative filtering), KNN with Z score, SVD, SVD++, NMF, Coclustering. All are implemented by surprise library with hyperparameter tuning. Final visualization made compare the results of all algorithms

Anime\_recommender3: Featured 1 simple deep learning recommender net.

Anime\_Clustering: Feature 4 clustering methods: K-means, Hierarchical Agglomerative Clustering, DBSCAN, Mean Shift.

Report: Describe goals, methods for this project and its finding.

Presentation: A PPT which used for presentation for this project.