

Recommending Your Favorite Manhwas(만화), Mangas, and Manhuas Using a Machine Learning Approach

Sharath Kannan and Eric Cho



What are Recommender Systems?

Three types (Isinkaye et. al, 2015)




- Content-based Filtering
 - Uses attributes of items to generate recommendations
- Collaborative Filtering
 - Combines item preferences from multiple users to generate recommendations
- Hybrid Filtering
 - Uses a combination of both strategies and algorithms to generate recommendations

Our Data

Uploaded on Kaggle by Victor Sorerio in 2021

- Information includes: title, description, rating, year, and tags

What are Mangas, Manhuas, and Manhwas (만화)?

- Comics from Japan 
- Comics from China/Taiwan 
- Comics from South Korea 

Data Analysis and Feature Engineering

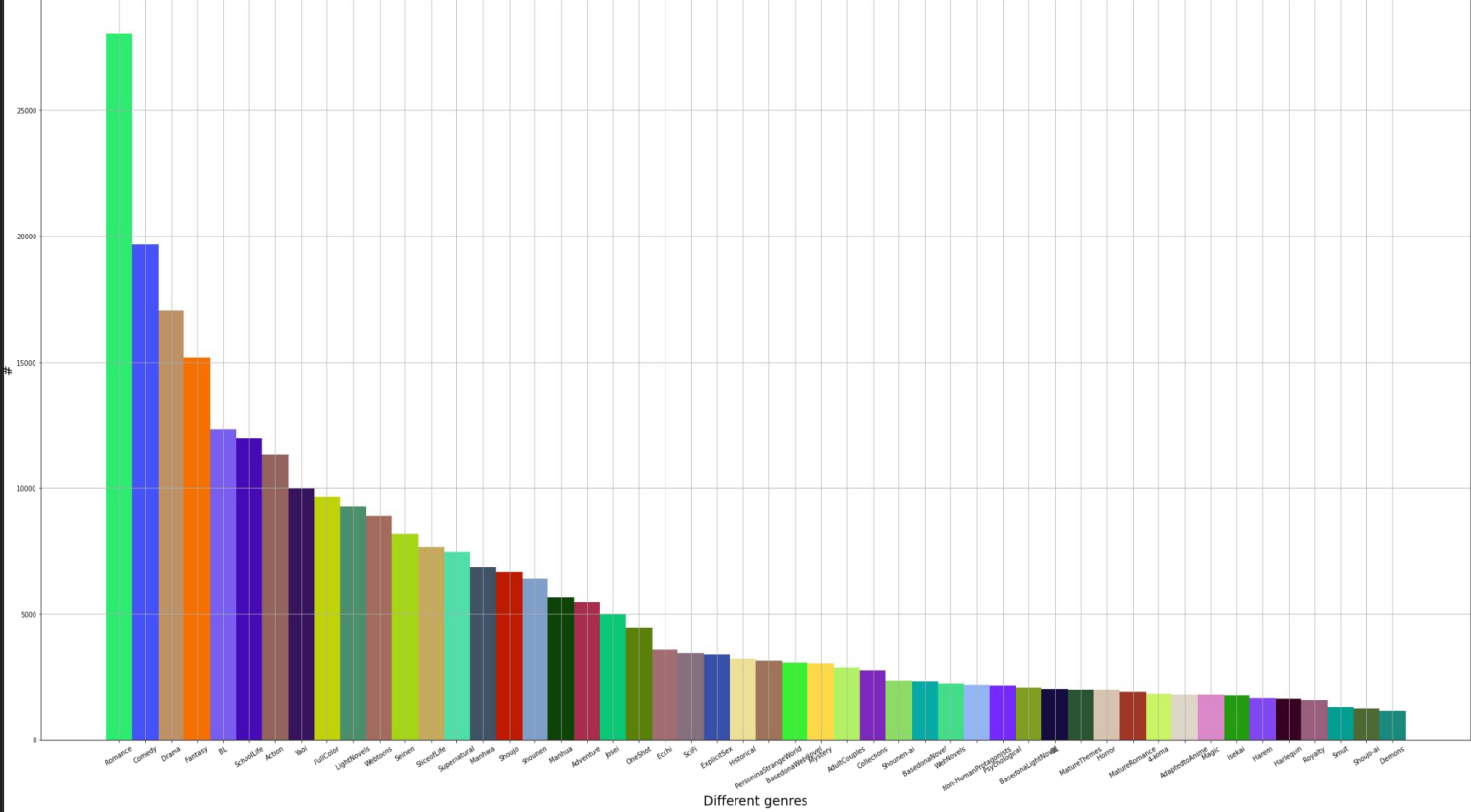
Data Analysis

- We discovered 70948 total manhwas,mangas and manhwas
- On the following graph on the next slide, you can see that the most popular genres are romance, action,comedy ,etc.
- We also filtered out NaN ratings, also we are recommending manhwas,mangas and manhwas from year 2000 and above

Feature Engineering

- “Tags” contains a list of columns - 568 Genres!
 - We are experimenting with the value of N, but it's currently 50
- One Hot Encode “tags” column

The different types of genre



K-Nearest Neighbors

Using Scipy implementation

- D=50? Curse of Dimensionality! Use Ball Tree:

A ball tree recursively divides the data into nodes defined by a centroid C and radius r , such that each point in the node lies within the hyper-sphere defined by r and C . The number of candidate points for a neighbor search is reduced through use of the *triangle inequality*:

$$|x + y| \leq |x| + |y|$$

X is our training data of one hot encoded values

Y is a sample row in the dataset

We find the K=10 best recommendations for Y.

Results from KNN (top 10 best recommendations)

Best Recommendations if You've Read: Demon Slayer: Kimetsu no Yaiba

['Action', 'Adventure', 'Comedy', 'Drama', 'Fantasy', 'Shounen', 'DeathofaLovedOne', 'Demons', 'Historical', 'MartialArts', 'Orphans', 'Siblings', 'ExplicitViolence', 'AdaptedtoAnime']

Akaboshi: Ibun Suikoden

['Action', 'Adventure', 'Comedy', 'Drama', 'Fantasy', 'Shounen', 'AncientChina', 'Historical']

Hero Tales

['Action', 'Adventure', 'Comedy', 'Drama', 'Fantasy', 'Shounen', 'MartialArts', 'AdaptedtoAnime']

Boruto

['Action', 'Adventure', 'Comedy', 'Drama', 'Fantasy', 'Shounen', 'Ninja', 'AdaptedtoAnime']

Black Cat

['Action', 'Adventure', 'Comedy', 'Drama', 'Fantasy', 'Shounen', 'Assassins', 'BountyHunters', 'Guns', 'Superpowers', 'AdaptedtoAnime']

Altair: A Record of Battles

['Action', 'Adventure', 'Drama', 'Fantasy', 'Shounen', 'Historical', 'MiddleEastern', 'Orphans', 'War', 'AdaptedtoAnime']

Peach Boy Riverside

['Action', 'Adventure', 'Drama', 'Fantasy', 'Shounen', 'DarkFantasy', 'Demons', 'AdaptedtoAnime']

Magi: The Adventures of Sinbad

['Action', 'Adventure', 'Comedy', 'Drama', 'Fantasy', 'Shounen', 'Dungeon', 'Genies', 'MiddleEastern', 'Orphans', 'OverpoweredMainCharacters', 'AdaptedtoAnime']

Toriko

['Action', 'Adventure', 'Comedy', 'Fantasy', 'Shounen', 'FoodandBeverage', 'Superpowers', 'AdaptedtoAnime']

Sengoku Strays

['Action', 'Adventure', 'Comedy', 'Drama', 'Fantasy', 'Shounen', 'FeudalJapan', 'Historical', 'Kendo', 'PersoninaStrangeWorld', 'TimeTravel']

Evaluation

There is no ground truth!!

So how do we evaluate this?

- Compare with other recommendation algorithms
- Change Approach: create a ground truth
 - Grab a subsample of items using KNN or manually—simulates a reading log
 - Label them with 1 (like) or 0 (dislike)
 - Have KNN (or another model) predict if an item is liked or disliked by the user.

Next Steps

- Implement the sample user reading list and scores
- Use a different model for recommendation
 - Decision Trees
 - Neural Network with Matrix Factorization
 - K-Means
- Experiment with different N values
- Add “rating” and “year” to KNN calculation