Recommendation System

By: Tommy Phung

Overview

Source: MovieLens

- **Authors:** GroupLens

Main Feature: Ratings

Goal: Recommendation system that suggested five movies based off prior users' ratings

Business Problem

- Over **8 million movies**
- Limited time
- Prone to suggestion
  - **70%** of Youtube content came from **recommendations**

Improve streaming services by offering **better recommendations**

Targeted Streaming Services

NETFLIX

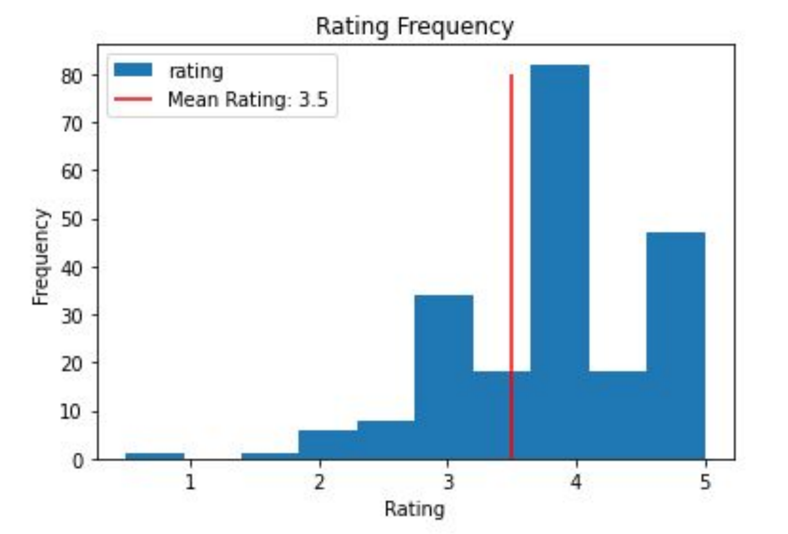
max

prime video

Data Understanding

Number of users: **610**
Number of Movies: **9,742**
Number of reviews: **100,836**
Dataset features: **Rating, Genre and Tags***
Rating Range: **0 (Bad) →5 (Good)**

- **0.5 steps**



Data Preparation

1. Remove **timestamp** from dataset (Not Needed)
2. Create **training** and **testing** set (Test Accuracy)
3. Standardize Genres (Content Filtering)

Methods

Main Library: **Surprise**
Memory Based Modeling:
KnnBasic, **KnnBaseline,** KnnWithMean

- **Clusters grouping similar items**

Model Based Modeling:

SVD Singular Value Decomposition

- Matrix decomposition

Best Performing: KnnBaseline

Additional Conditions

1. Model won’t recommend seen movies
2. Offer new users top movies (Cold Start)
3. New Provide recommendation based off movie provided

Cold Start

A cold start is a problem when a **new user or item** is added without **prior history** in the current system.

User Case: A new user is added needs a recommendation.

Item Case: A new movie is added and need to be recommended.

Ex. New User Recommendations

Most Rated Movies	Most Popular Movies
Paper Birds (Pájaros de papel) (2010)	Pulp Fiction (1994)
Act of Killing, The (2012)	Shawshank Redemption, The (1994)
Jump In! (2007)	Forrest Gump (1994)
Human (2015)	Silence of the Lambs, The (1991)
L.A. Slasher (2015)	Matrix, The (1999)

Most Rated: All rated 5 stars
Most Popular: Movies with the most users rated.

New Movies

1. Average Action Movie Rating: ~3
   - Never be recommended…
2. User Favorite Genre weight?

<u>User 1 Top Watched Genres</u> → Recommend Action Movies

action|adventure|sci-fi	11
comedy	11
action|adventure|thriller	8
action|drama|war	8
comedy|drama	6

Conclusion

Best Performing Model: **KnnBaseline (.8708)**
Cold Starts: Content Filtering and other method are needed for new users and items.

Two filtering are needed for a recommendation system

- ➢ Collaborative for current users and movies
- ➢ Content for new users and movies

Limitations

21 users → 26,663 of the total reviews (25%)

- Bias toward certain genres, series, age, etc.

5 - 10% write reviews unprompted

- More likely to leave negative reviews if any.

Review Bombing

- A group of people leaving a large amount of negative reviews to negatively impact the movies

Next Steps

Factor Genres

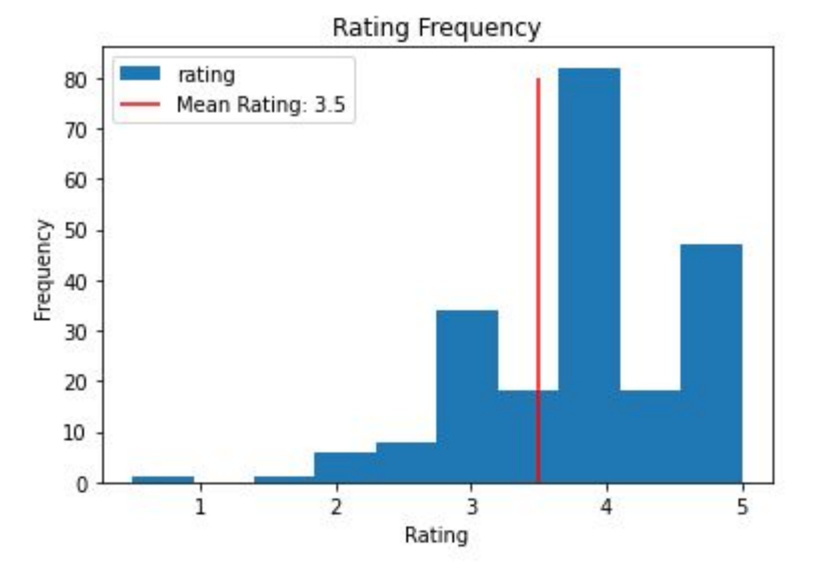
- Certain combination might be **favorable**

More Diverse Ratings

- **Bell Curve Distribution**

More Data

- More data = Better modeling



Questions?

Original Source:
https://grouplens.org/datasets/movielens/latest/
Email: phungtommy109@gmail.com
Github: https://github.com/Tommyphung1
Notebook: https://github.com/Tommyphung1/Project_4