A Recommendation System for Movies

BY: ALEC HING

Outline

- Overview
- Data Involved
- Approach
- Methods
- Results
- Improvements
- Conclusion

Overview

- A dataset from MovieLens research lab that consists of about 100,000 user ratings was analyzed
- Through several machine learning techniques, movie recommendations were made for these users
- Recommendations were also made for new users with no prior information

Data Involved

- The datasets used in this project are from the MovieLens research lab at the University of Minnesota
- Main categories analyzed were Movie Id, Title, Genre, User Id, and Rating
- There were 9,724 different movies, 610 different users, and a total of 100,836 movies rated
- Ratings were on a scale of 0 to 5 with the lowest rating being 0.5 and the highest 5.0

Approach

- The most important information that needed to be investigated were:
  - What users rated and what they didn’t
  - How many movies and users are there in this data
  - What was the lowest & highest ratings movies got
  - What movies received the most ratings
- 2 systems will be used to create a proper recommendation system
  - First – Collaborative Filtering: Users with similar interests will probably like the same thing
  - Second - Content based filtering: If a user likes an item, they may like similar items

Methods

- There were many methods to consider that resulted in a better or worse accuracy of the predictions being made
- To name a few:
  - The ALS and matrix factorization models were tested but these tend to show better results in massive dataset – 100,000 is considered “small”
  - Memory based methods like the K Nearest Neighbor (KNN) variations proved to have a better accuracy – These algorithms are used for its simplicity
- Because this is a “smaller” dataset, computing time was not a priority

Results

- Out of the 6 methods used, the most accurate one was KNNBaseline
- This is a more advanced model than the rest that accounts for other variables
- Its accuracy was calculated to be about 0.55 off while the rest were around 0.80

Improvements

- Fine tuning the parameters of the models used can help improve accuracy
- Offer a new user more options to choose from instead of just from the top ten rated movies
- Take into consideration how many people rated a movie

Conclusion

- The movies with the most ratings tended to also be some of the most popular
- Just because a movie got a 5.0 does not mean everyone thinks that
- Multiple people who feel a certain way about a movie show the truth
- New users need to give a baseline of where their interests are to make appropriate recommendations

Thank You

- Email: ahing619@gmail.com
- GitHub: @ahing
- LinkedIn: https://www.linkedin.com/in/alec-hing