Project 2: Movie Recommendation System

# **Lara Clasen Spring Term 2020 LLClasen.github.io**

**Week 6 Check-In**

**Updated Proposal**

# Which Domain?

*What domain is this data going to come from? Please list 10 references for your domain (with a brief annotation) to use to make sense of what you’re doing with these data.*

My data set for Project 2 comes from the Grouplens organization which pulled data stemming from the movie recommendation system MovieLens.

References

1. Rocca, B. (2019, June 12). Introduction to Recommender Systems. Retrieved from <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada>
   1. Offers a detailed overview of the different recommendation systems that are available and in what situations you might use them.
2. Poudyal, R. (2018, June 27). Latent Factor based method in collaborative filtering. Retrieved from <https://medium.com/@rabinpoudyal1995/latent-factor-based-method-in-collaborative-filtering-77756a02f675>
   1. This article reviews just one type of recommendation system, and uses generalized examples to explain how it works.
3. Das, S. (2015, September 24). Beginners Guide to learn about Content Based Recommender Engine. Retrieved from <https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/>
   1. This article reviews content-based recommender engines popular services such as Facebook to demonstrate it usage.
4. Kirzhner, E. (2018, May 11). Machine Learning. Explanation of Collaborative Filtering vs Content Based Filtering. Retrieved from <https://codeburst.io/explanation-of-recommender-systems-in-information-retrieval-13077e1d916c>
   1. This article is very helpful for my case as it takes the time to compare two types of more complex recommendation algorithms.
5. Ansari, A., & Kohli, R. (2000, August). Internet recommendation systems. Retrieved from [https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/385/Internet Recommendation Systems.pdf](https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/385/Internet%20Recommendation%20Systems.pdf)
   1. An older article, this write-up offers some more numerical insight into the recommendation system.
6. Chauhan, N. S. (2019, July 9). Implement K-Nearest Neighbors classification Algorithm. Retrieved from <https://towardsdatascience.com/implement-k-nearest-neighbors-classification-algorithm-c99be8f14052>
   1. This article offers a valuable example of using the K-Nearest Neighbors classification algorithm.
7. Recommender Systems. (n.d.). Retrieved from <http://recommender-systems.org/>
   1. A very basic yet useful resource for understanding different recommendation system types.
8. Liao, K. (2018, November 19). Prototyping a Recommender System Step by Step Part 1: KNN Item-Based Collaborative Filtering. Retrieved from <https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-1-knn-item-based-collaborative-filtering-637969614ea>
   1. Another helpful resource for the utilization of the K-Nearest Neighbors algorithm, specifically for use with collaborative filtering.
9. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (n.d.). Item-Based Collaborative Filtering Recommendation Algorithms. Retrieved from <http://files.grouplens.org/papers/www10_sarwar.pdf>
   1. This paper discusses item to item collaborative filtering specifically.
10. Ricci, F. (n.d.). Item-to-Item Collaborative Filtering and Matrix Factorization. Retrieved from <https://www.ics.uci.edu/~welling/teaching/CS77Bwinter12/presentations/course_Ricci/13-Item-to-Item-Matrix-CF.pdf>
    1. An interesting presentation presenting item to item collaborative filtering in comparison with other similar methods.

# Which Data?

*What is the dataset you’ll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.*

The data set was originally obtained from the URL: <https://grouplens.org/datasets/movielens/latest/>. This group out of the University of Minnesota specializes in recommender systems and other online computing systems and researches their use and abilities. The data set contains rating and tagging activity from MovieLens, and consists of 105,339 ratings and 6,138 tagging applications across 10,329 movies. Users are represented simply by an ID without any other demographic detail. All users had rated at least 20 movies to be selected. README available here: <http://files.grouplens.org/datasets/movielens/ml-latest-small-README.html>

# Research Questions? Benefits? Why analyze these data?

*How are you proposing to analyze this dataset? This is about your approach. Here, you’ll be proposing your research questions as well as justifications for why you’d offer these data in this way.*

With this data set I plan to build a movie recommender system. Recommendation systems are algorithms that work to offer relevant suggestions to the user based on a particular industry or product area. This data is important to analyze because it combines two huge pieces of the entertainment experience in today’s world: recommendations and movie information. It will allow users to obtain accurate recommendations for movie choices without the requirement of doing anything but looking at their phone or using the app which utilizes such an algorithm.

# What Method?

*What methods will you be using? What will those methods provide in terms of analysis? How is this useful?*

There are a few different types of recommendation systems that can be used for this sort of data. A simple recommender is pretty basic and offer suggestions based on what the general public “likes”. However, I’d like to get a bit deeper and so I want to build either a content-based or collaborative filtering recommendation system for this movie data. Both of these use the specific user’s behavior to make suggestions. While content-based recommenders base their suggestions directly on user preferences for product features, collaborative filtering (as the name suggests) makes assumptions based on the user’s preferences as they relate to other user preferences. Because we have quite a bit of “content” to work with in this data set, we should be able to build either type of algorithm. I plan to build an item to item collaborative filtering system.

# Potential Issues?

*What challenges do you anticipate having? What could cause this project to go off schedule?*

Since we have worked with k-NN algorithms more than once, I don’t anticipate having much trouble in that respect. However, I don’t believe that I have experience with an item-based recommender and so I may run into some unanticipated trouble there. My main goal is to create a collaborative filtering algorithm so that is what I will focus on.

# Concluding Remarks

*Tie it all together. Think of this section as your final report’s abstract.*

In this second project I plan to build a recommendation system to predict a user’s preference in movie choice most effectively. I will use MovieLens metadata to create a recommender that can work off of user behavior and movie characteristics to produce suggestions relevant to the user. This project and its resulting algorithm will be incredibly useful in a world where many people rely on reviews and recommendations before they will try things, especially new movies. By offering accurate recommendations, any application utilizing the algorithm offers a valuable service to their user.

**Week 6 Check-In Questions**

**Any surprises from your domain from these data?**

I did not really experience any surprised, but I changed the scope of my project a bit since my initial proposal last week. After beginning with the restaurant recommender project, I came across movie recommender data and found it to be much more interesting to work with. This data has provided a more fulfilling project so far.

**The dataset is what you thought it was?**

Yes, my movie recommendation data set has been basically what I expected to this point. The data source provided a good amount of background on the data so I was pretty familiar with what I would find from the get-go.

**Have you had to adjust your approach or research questions?**

Not yet; I am still working toward an item-based collaborative filtering system using these data. I have not run into any issues with this goal and it seems to be working well with this type of data set.

**Is your method working?**

Yes! I am pleased to be successfully building the type of recommender system that I had initially set out to build with this data set.

**What challenges are you having?**

Currently I am having to determine the best visualizations to use to present the data that I am working with. However, I believe that my data visualization course prepared me well for this and I am looking forward to finding solutions. I am especially excited to be working in R for this project and creating visualizations using R packages.