# Personalized Movie Recommendation System

Raj Kamal Yadav, Rajat Bansal, Manish Mahalwal

2016076, 2016260, 2016054

## Introduction

In today's world of Netflix and Chill, recommendation systems have become an essential part of our life, especially due to the increasing of choices available to us. For a media commodity like movies, suggestions are made to users by finding user profiles of individuals with similar tastes. We shall be using both Content-Based approaches with features such as cast, crew etc. as well as Collaborative Filtering for which the input to our algorithm will be observed user ratings in the past to recommend items of interest to users. For baselines, we implement Matrix Factorisation and TF-IDF approaches. We also implement BPR for personalised ranking of movies for each user. This technique proves to be far better than the two baselines MF and TF-IDF.

## Literature Review

Two main types of recommender systems are:

- **Content-based:** It recommends movies on the basis of movie-movie similarity. Useful for cold-start approach.
- **Collaborative Filtering**: It looks for various patterns in the activities of users and generates user targeted recommendations. Useful for personalised recommendations.

In Movie recommendation, we predict a personalized ranking on a set of movies. There are many methods for this task like matrix factorization and or adaptive k-nearest-neighbour(kNN). Although these techniques are designed for movie recommendation task, we can't use them for ranking. We will implement a specialized optimization technique called BPR-OPT for personalized ranking. We use it maximize the posterior probability which is obtained from a Bayesian analysis of the problem. The learning part of the problem is based on bootstrap sampling and stochastic gradient descent. Bayesian Personalized Ranking is used to create a user-specific ranking for a set of movies.

- **BPR-OPT:** implement the generic criterion for BPR-OPT using the maximum posterior estimator for optimum personalized ranking of items. This optimization is to related to maximizing the area under curve of an ROC.
- **LearnBPR:** implement a learning algorithm based on stochastic gradient descent. This technique is superior than std grad descent techniques used for optimization of BPR-OPT.

## Dataset Description

Dataset[1] used is available at Kaggle[1] however it contains 26 million ratings from 270,000 users and 45,000 movies. We use a subset of the dataset containing 9,000 movies and 100,000 ratings over 700 users. The data is provided to us in the form of:

- **Credits.csv**: About movie's cast and crew information
- **Keywords.csv**: Keywords associated to each movie
- **Links.csv**: IMDB and TMDB IDs' for all movies
- **Rating.csv**: Contains 100,000 ratings from 700 users to 9,000 movies

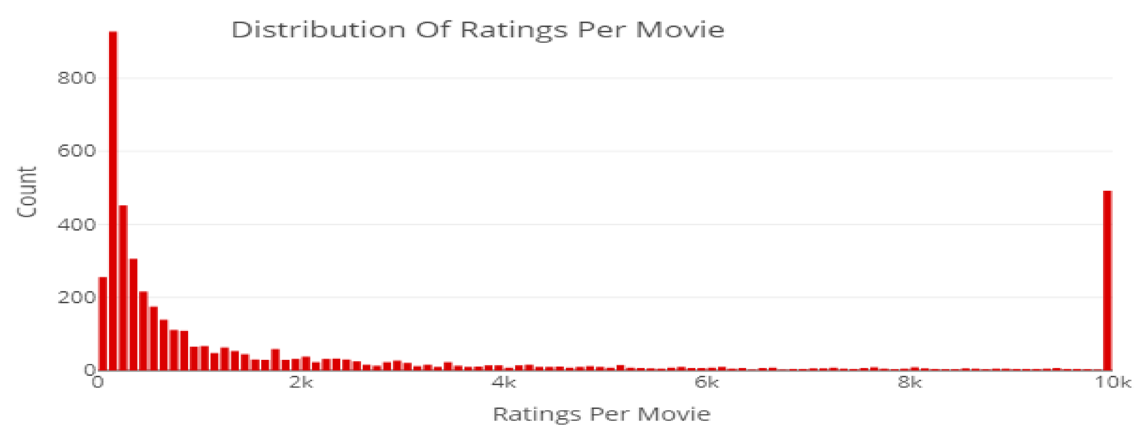


Figure 1: Distribution of Ratings per movie

## Implemented Baselines

- TF-IDF Similarity[2]
- User-Based Collaborative Filtering

TF-IDF exploits the common assumption that **similar movies share similar keywords and cast**. Pairwise Cosine Similarity between all movies is calculated (using Keywords and Cast,Crew information) and their weighted combination is stored in a final similarity matrix.

Collaborative Filtering method is based on the simple premise that **similar users share similar interest**.

## Bayesian Personalized Ranking

The algorithm[3] assumes that the user has interacted with the content in some manner such as – number of clicks, purchases, number of views. We have used the ratings from the users as explicit feedback to the algorithm.

It uses a pair-wise item feedback matrix to reconstruct a personalized total ranking for each user by transforming a positive only feedback into positive and negative feedback in terms of pairs of items (i,j), where the user prefers i over j (positive) and correspondingly rephrased dislikes j over i(negative).

## Result Compilation

We propose the following architecture for evaluating TF-IDF approach as discussed:

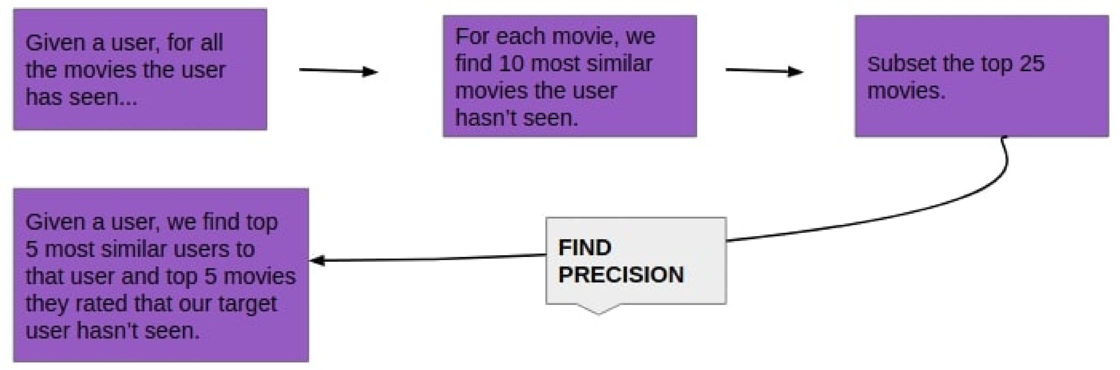


Figure 2: Evaluation Architecture

The intuition behind the above architecture is that the movies similar to those the user have seen should also be seen by similar users and liked by them.

We obtained a **precision of 36.84%** on User3 while testing the above architecture.

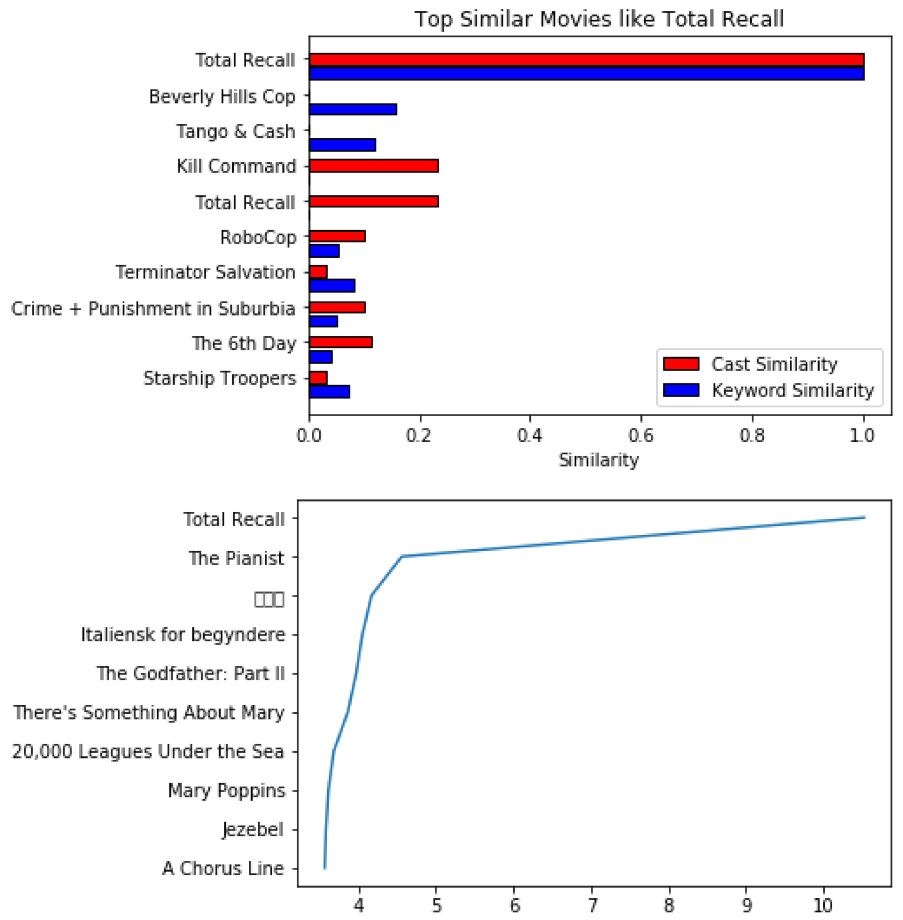


Figure 3: TF-IDF v/s BPR for movie

## Conclusion

After having fair understanding of each of the three recommendation algorithms, we observed that Text Frequency and Inverse Document Frequency though provides reasonable suggestions but does not account user opinion in it. On the other hand User Based Collaborative Filtering does extract users that are similar to the user of interest and recommend their movies that the user of interest has not seen yet.

## References

[1] Rounak Banik.
The movies dataset.
In *https://www.kaggle.com/rounakbanik/the-movies-dataset/version/7ratings_small.csv*.

[2] Tianyi Liu Shujia Liang, Lily Liu.
Personalize movie recommendation system.
In *http://cs229.stanford.edu/proj2018/report/128.pdf*.

[3] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme.
Bpr: Bayesian personalized ranking from implicit feedback.
In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press, 2009.