

Final Project

LEFOGLOU POLYDEFKIS

1 Introduction

A recommendation system is a type of software application or algorithm designed to suggest items or content to users. The primary goal of recommendation systems is to improve user experience by providing personalized and targeted suggestions. These systems are widely used in various online platforms to help users discover products, services, or content that they might find interesting or relevant. There are generally two main types of recommendation systems:

1) Collaborative Filtering: In this approach, the system analyzes information about preferences, behaviour, and activities of all users to predict what you might like. Simply put, the system recommends items that other users with similar tastes and behaviour liked. The main assumption of this method is that people who liked similar products in the past will also like similar products in the future.

2) Content-Based Filtering: These methods are based on product descriptions and user preferences. This type of system recommends products similar to the products the user has liked in the past.

Specifically, this project is focused on movie recommendation systems. Movie recommendation systems try to suggest movies for users based on their previous viewing behavior and preferences. These systems analyze data such as users' ratings, reviews, and viewing histories to generate personalized recommendations.

2 Dataset and Features

For this project we used the MovieLens Latest Dataset, which is recommended for education and development (<https://grouplens.org/datasets/movielens/>). It contains 100836 ratings and 3683 tag applications applied to 9742 movies by 610 users.

userid	movieid	rating
0	1	1
1	1	3
2	1	6
3	1	47
4	1	50
...
100831	610	166534
100832	610	168248
100833	610	168250
100834	610	168252
100835	610	170875

Figure 1: Ratings.

userid	movieid	tag	timestamp
0	2	60756	funny
1	2	60756	Highly quotable
2	2	60756	will ferrell
3	2	89774	Boxing story
4	2	89774	MMA
...
3678	606	7382	for katie
3679	606	7936	austere
3680	610	3265	gun fu
3681	610	3265	heroic bloodshed
3682	610	168248	Heroic Bloodshed

Figure 2: Tags.

movieid	title	genres
0	1	Toy Story (1995)
1	2	Junany (1995)
2	3	Grumpier Old Men (1995)
3	4	Waiting to Exhale (1995)
4	5	Father of the Bride Part II (1995)
...
9737	193581	Black Butler: Book of the Atlantic (2017)
9738	193583	No Game No Life: Zero (2017)
9739	193585	Flint (2017)
9740	193587	Bungo Stray Dogs: Dead Apple (2018)
9741	193609	Andrew Dice Clay: Dice Rules (1991)

Figure 3: Movies

3 Methods

In this project we explored various methods to create a movie recommendation system. The first method is Collaborative Filtering with Singular Value Decomposition(SVD). SVD is a matrix factorization approach. That means that it tries to find two (or more) matrices such that when you multiply them you will get back the original matrix. In our case the original matrix R is the ratings matrix, which is a sparse matrix (the users haven't rated every movie). Additionally each user and movie is described with F latent features. The original matrix is then decomposed in the product of the user factors matrix P and the movie factors matrix Q . Using P and Q we predict each user's rating for every movie and we create a full ratings matrix. Specifically, the prediction for user 'u' about movie 'm' is:

$$p_{u,m} = q_m^T p_u = \sum_{f=1}^F q_{mf} * p_{uf} \quad (1)$$

Then we used a Content Based Filtering approach. For this approach we had to create a movie-feature matrix, which is a matrix consisting of binary values, 1 if the movie has that feature and 0 if it doesn't. Then by using cosine similarity (which is a value that is bound by a constrained range of 0 and 1) we found for each movie other movies similar to it (if cosine similarity is close to 1 it means that the movies are similar to each other). We used this method twice, once with the feature being the movie genre and once with the feature being a combination of the genre and the tag. Finally we used neural networks. The inputs were binary values that show if a movie belongs to a certain genre or not and the output was a specific user's ratings.

4 Experiments/Results

After we implemented the methods above we experimented with them to observe the results. The Content Based approach requires the user to input a favorite movie to recommend similar movies. Comparing the top 5 results (either with the feature being the genre or genre/tag) to the top 5 recommendations that IMDb makes we observe that 1-2 out of 5 recommendations match. That isn't a great result but that's expected because the genres/tags aren't enough to make an accurate recommendation. We probably need more data for the movies like actors or director. As for the Collaborative Filtering

method, it is hard to know how accurate our recommendations are since they are tailored to a specific user. The one observation we could make is that the recommendations our system provides are usually movies that are considered popular probably because they were rated highly from a lot of the users in our database.

5 Conclusion

In conclusion the algorithms we used in this project provided adequate recommendations given the dataset we had. Nevertheless, with a dataset that provides more information about the users and the movies they could definitely be improved in order to make more accurate recommendations.