ReelRx

A Movie Recommendation App Powered by Machine Learning

Group 3: Hoa Roach, Jammy Lo, Tim Schurmann, Tiffany Burns

Project Overview

ReelRx 

ReelRx is a movie recommendation app that allows a user to input a movie, select whether to receive obscure recommendations, and receive a list of movie recommendations based on the input.

Our website offers a dashboard that allows users to explore top movies by applying filters for age, gender, and even occupation and a table to explore database movies by genre.

Inspiration

Reel Rx

ReelRx

Northwestern Medicine®

Roku

GoodRx

GoodRx

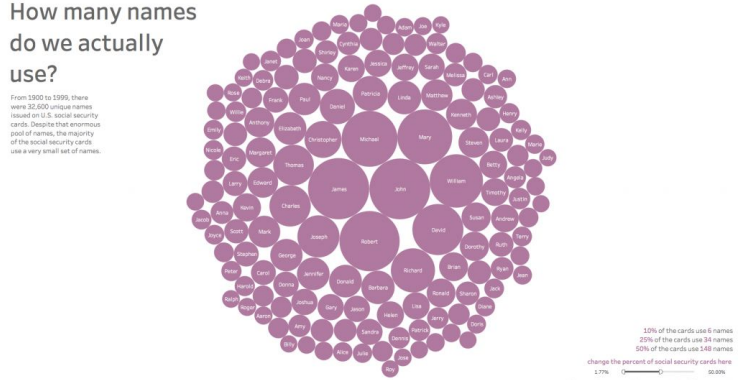
#281D0E

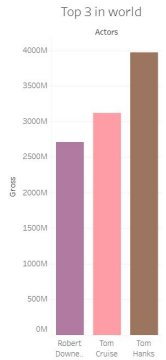
#5925BD

#AA6AFA

#CEC3FE

#EDF3AE





Where does the data come from?

This dataset contains a set of movie ratings from the MovieLens website, a movie recommendation service. This dataset was collected and maintained by GroupLens, a research group at the University of Minnesota. There are 5 versions included: "25m", "latest-small", "100k", "1m", "20m". In all datasets, the movies data and ratings data are joined on "movieId". The 25m dataset, latest-small dataset, and 20m dataset contain only movie data and rating data. The 1m dataset and 100k dataset contain demographic data in addition to movie and rating data.

For this reason, we chose to use the “1M” dataset as it is the largest MovieLens dataset that contains demographic data.

We wanted more user data to be able to give more specific movie recommendations.

Visit ReelRx

How does ReelRx work?

ReelRx is built on a collaborative, item based filtering model. It uses a nearest neighbors model with cosine similarity to predict items (movies with ratings) a user may be interested in by similarity to decisions (ratings applied to movies) made by other users. ReelRx takes two inputs: A movie name and whether or not to receive obscure recommendations.

The backbone of ReelRx is a movie / user ratings matrix:

```
final_dataset = ratings.pivot(index='movie_id',columns='user_id',values='rating')
```

```
final_dataset.head()
```

user_id	1	2	3	4	5	6	7	8	9	10	...	6031	6032	6033	6034	6035	6036	6037	6038	6039	6040
movie_id																					
1	5.0	0.0	0.0	0.0	0.0	4.0	0.0	4.0	5.0	5.0	...	0.0	4.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	3.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	...	0.0	0.0	0.0	0.0	2.0	2.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0

5 rows × 6040 columns

What is an “obscure” movie?

An analysis was done to derive a threshold of 20th percentile to distinguish between obscure and not obscure movies. On this dataset, this means any movie with between 1 and 23 total ratings is considered “obscure.”

```
no_user_voted = ratings.groupby('movie_id')['rating'].agg('count')
```

```
no_user_voted.describe()
```

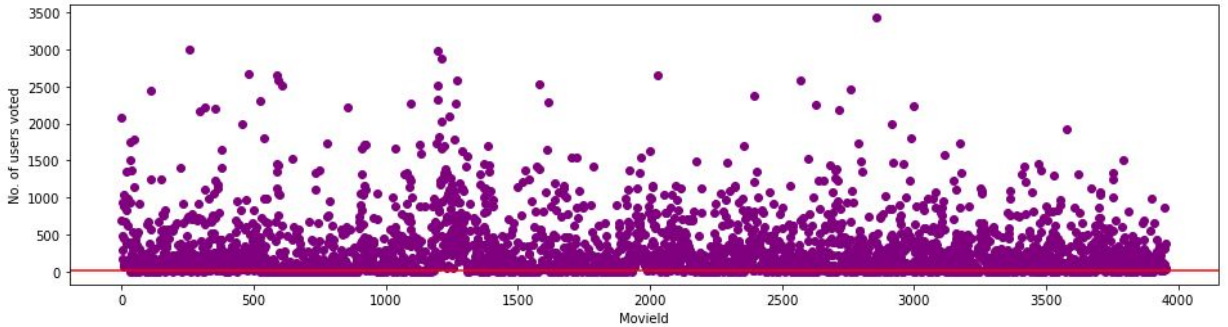
```
count    3706.000000
mean      269.889099
std       384.047838
min         1.000000
25%        33.000000
50%       123.500000
75%       350.000000
max      3428.000000
Name: rating, dtype: float64
```

```
ptile = 20
p = np.percentile(no_user_voted,ptile)
print(p)
print(len(no_user_voted)*(ptile/100))
```

```
23.0
741.2
```

```
f,ax = plt.subplots(1,1,figsize=(16,4))

plt.scatter(no_user_voted.index,no_user_voted,color="purple")
plt.axhline(y=23,color='r')
plt.xlabel('MovieId')
plt.ylabel('No. of users voted')
plt.show()
```

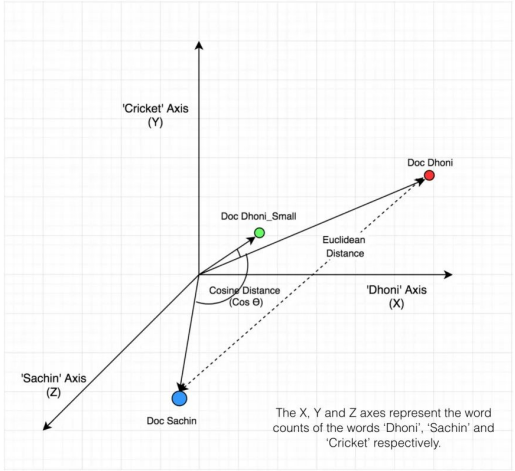


How does ReelRx work? (cont'd)

Once a dataset is selected (obscure or not obscure) and a movie is entered, ReelRx applies a cosine similarity nearest neighbors model.

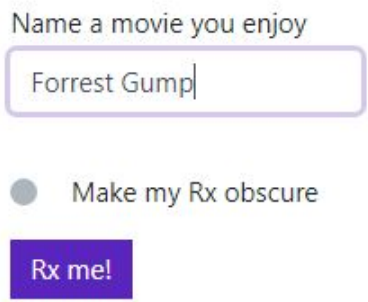
Cosine similarity is a measure of directional or angular similarity, as opposed to a measure of positional distance.

Projection of Documents in 3D Space



This method gives qualitative preference to movies rated rather than quantitative preference to numbers of ratings applied.

Thus, with two simple inputs, a user gets a list of movie recommendations!



Rx Score	*Click to learn more*
0.5077554472	Apollo 13 (1995)
0.5062098547	Ferris Bueller's Day Off (1986)
0.5043727132	Dave (1993)
0.5021588993	Wedding Singer, The (1998)
0.5018734205	When Harry Met Sally... (1989)

Boundaries & Limitations

- Dataset only includes movies made in 2000 or earlier
- 75 % of users are Male
- 40% of users are between ages 25-34
- 13% of users selected “other / not specified” for occupation
- 43% of movies in this database were released between 1990 - 1999
- Movie release dates end in 2000 (21 years of movie data missing)

Future Work

- Gather a larger data set that includes more user information and is a more wide spread sample across age, gender, and occupation

- Add more relevant movies and connect to an API to allow new releases monthly

- Add the ability to create a user profile that would allow users to save their submissions and get better results

- Connect to / Launch a streaming service, allowing users the ability to select the recommended movies to watch

References

Data: https://grouplens.org/datasets/movielens/

Model: https://www.kaggle.com/johnwill225/movie-recommendations

API: http://www.themoviedb.org

Cosine Similarity: https://www.machinelearningplus.com/nlp/cosine-similarity/#2whatiscosinesimilarityandwhyisitadvantageous

Questions?