A Recommendation System for Movies

BY: ALEC HING

Outline

- Overview

- Data Involved

- Approach

- Methods

- Results

- Improvements

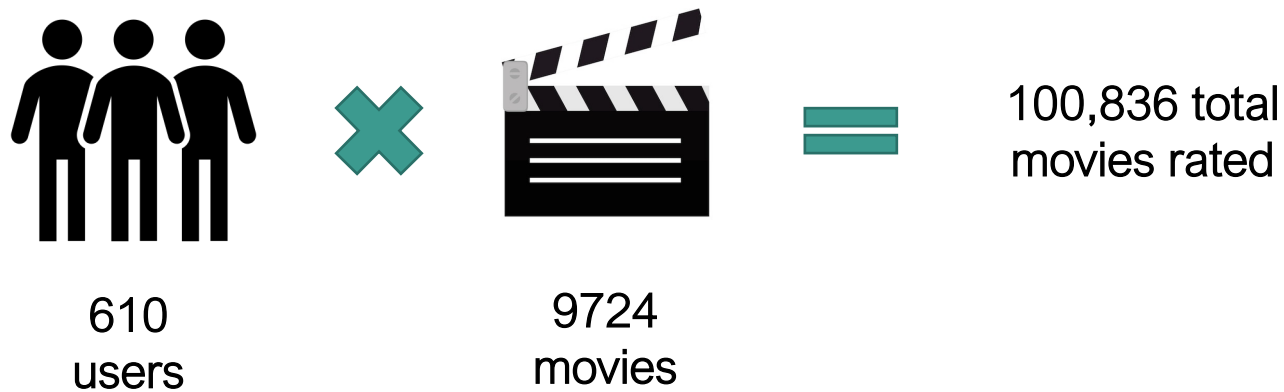- Recommendations

- Conclusion

Overview

- Provide Netflix user's meaningful recommendations they may like

- How do we know what movies a user may like based on prior likes?
  - What about no 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

- Ratings were on a scale of 0 to 5

610 users ✖ 9724 movies = 100,836 total movies rated
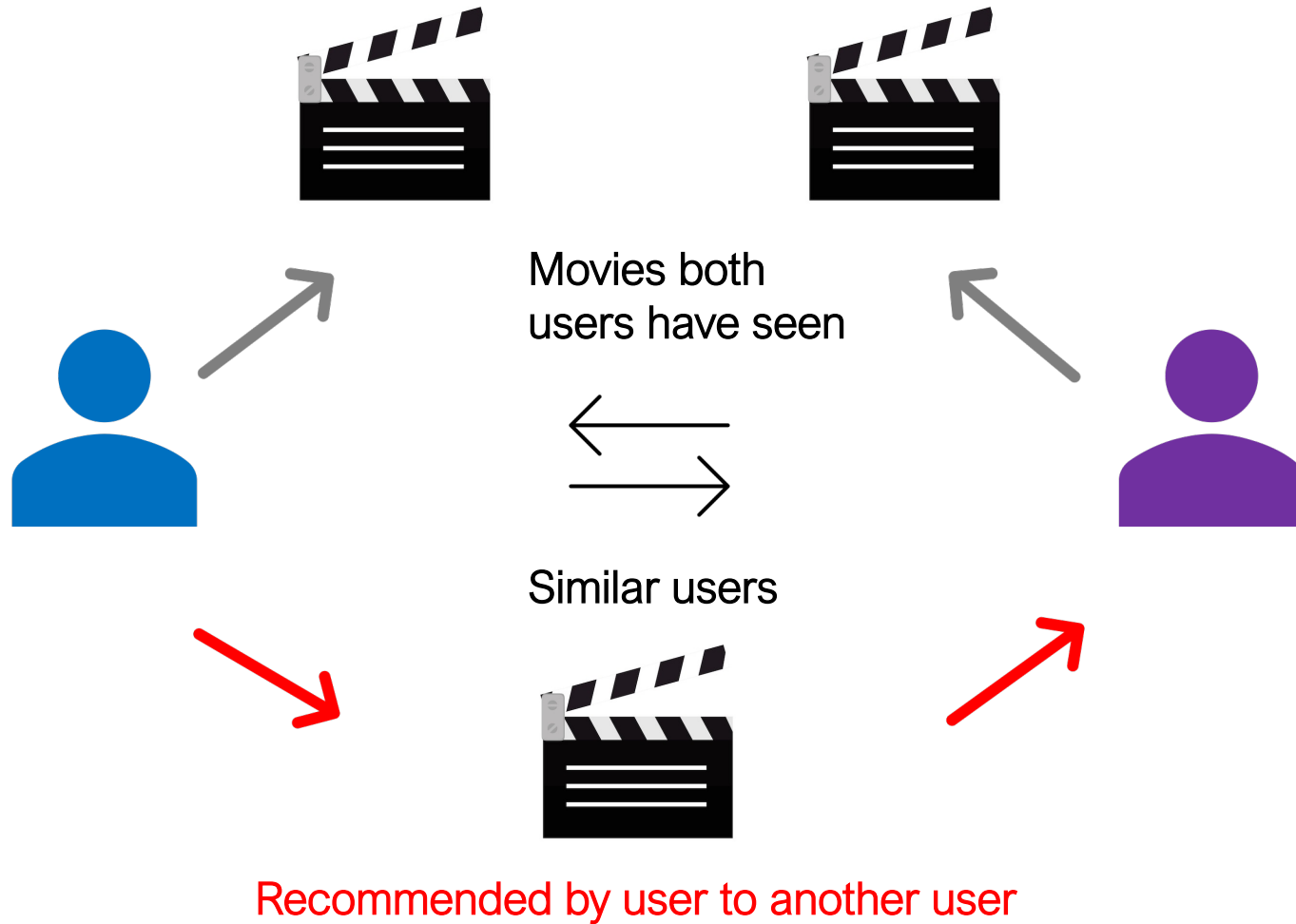
Approach

- Questions:
    - What have they not seen?
    - Quantity of movies
    - Total ratings per movie
- 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
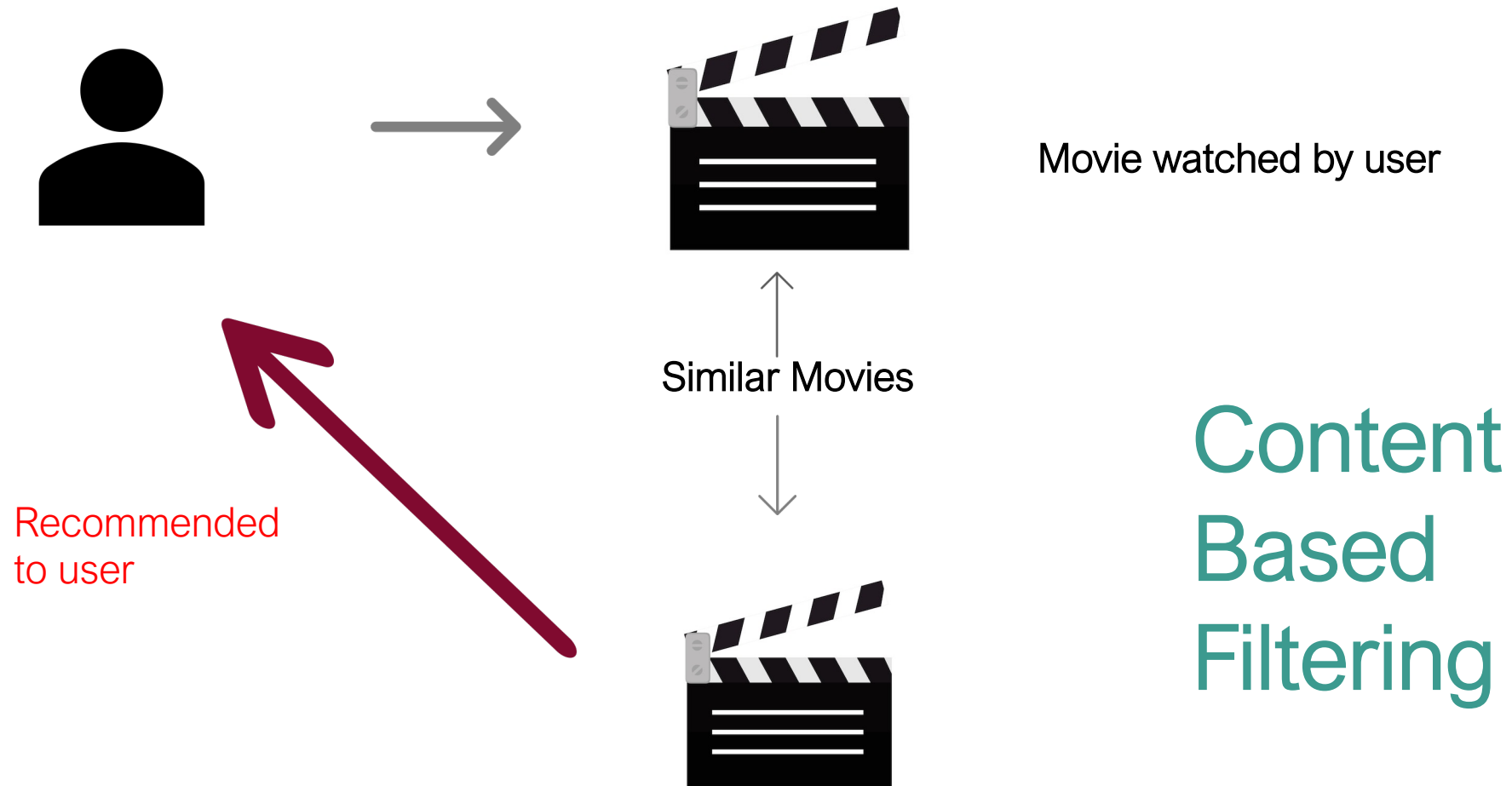
Methods

- Models:
    - The ALS and matrix factorization models used for massive datasets
    - Memory based methods used for simplicity

- Because this is a "*smaller*" dataset, computing time was *not* a priority

Methods – Collaborative Filtering

Movies both
users have seen

Similar users

Recommended by user to another user

Methods – Solution to Cold start Problem

Movie watched by user

Similar Movies

Recommended to user

Content Based Filtering

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

- Ask about Genres, Length, Actors

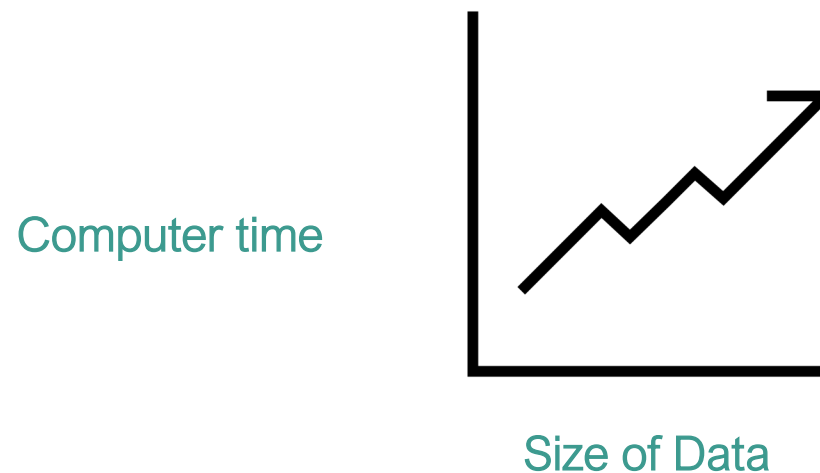- Number of ratings per movie (See next slide)

Improvements

title	rating
Gena the Crocodile (1969)	5.0
True Stories (1986)	5.0
Cosmic Scrat-tastrophe (2015)	5.0
Love and Pigeons (1985)	5.0
Red Sorghum (Hong gao liang) (1987)	5.0
...	...
Don't Look Now (1973)	0.5
Journey 2: The Mysterious Island (2012)	0.5
Joe Dirt 2: Beautiful Loser (2015)	0.5
Jesus Christ Vampire Hunter (2001)	0.5
Fullmetal Alchemist 2018 (2017)	0.5

VS

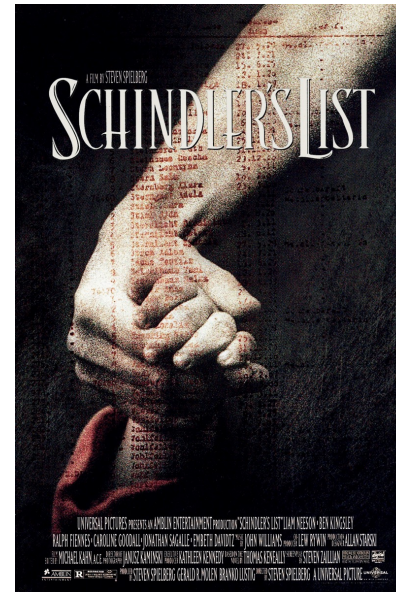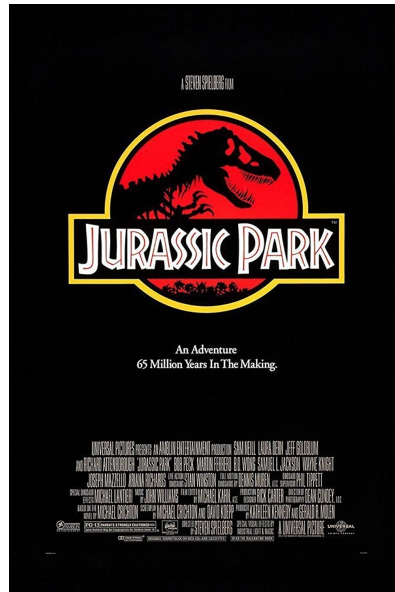| title	rating	Total Ratings
Forrest Gump (1994)	4.164134	329
Shawshank Redemption, The (1994)	4.429022	317
Pulp Fiction (1994)	4.197068	307
Silence of the Lambs, The (1991)	4.161290	279
Matrix, The (1999)	4.192446	278
Star Wars: Episode IV - A New Hope (1977)	4.231076	251
Jurassic Park (1993)	3.750000	238
Braveheart (1995)	4.031646	237
Terminator 2: Judgment Day (1991)	3.970982	224
Schindler's List (1993)	4.225000	220

Recommendations

- Compute time – Small dataset vs. <u>Big dataset</u>

- Recommend movies similar users would like

- Ask users about their interests

- Consider *number* of ratings

Computer time

Size of Data

Conclusion

- Most rated -> Some of the most popular

- More ratings for a movie = More accurate

- Need a baseline of user's interests

Thank You

- Email: ahing619@gmail.com

- GitHub: @ahing

- LinkedIn: https://www.linkedin.com/in/alec-hing