UCLA

CS145 Project Introduction

Movie Rating Predictions

Instructor: Yizhou Sun

TAs: Yunsheng Bai, Shengming Zhang

01/14/2019

UCLA

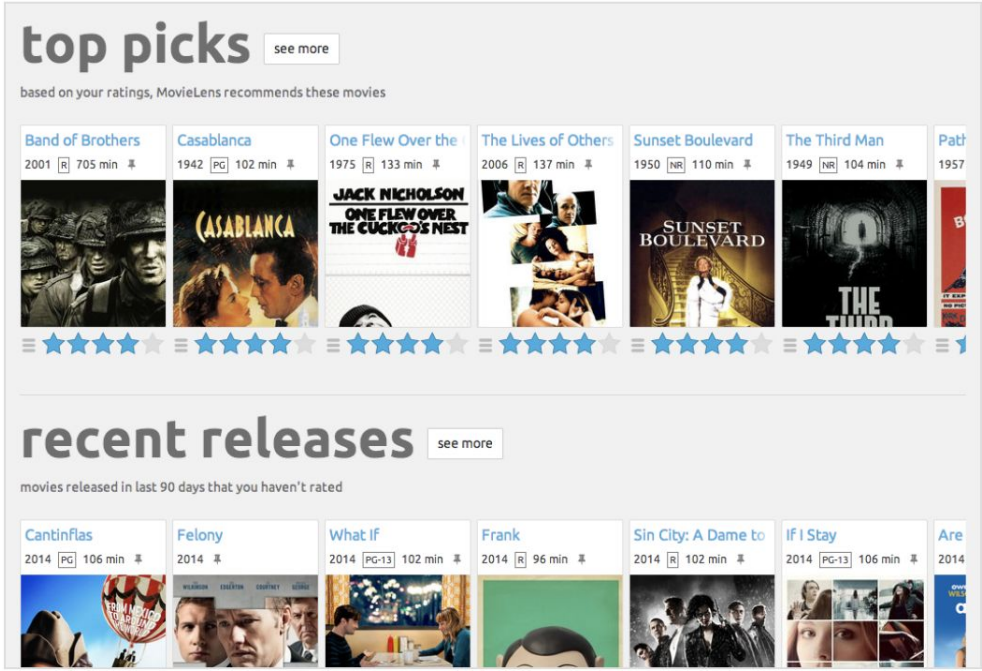
Project Introduction

- Background & Motivation
- Project Task
- Dataset
- Evaluation
- Project Deadlines and Grading

Background

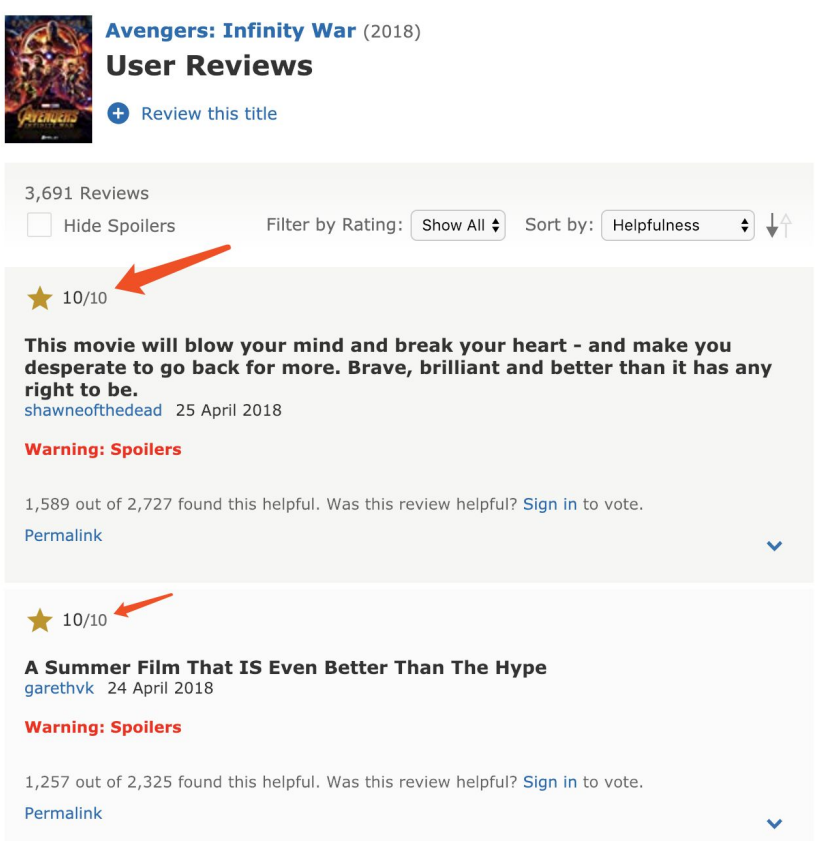
MovieLens Dataset

MovieLens helps you find movies you will like. Rate movies to build a custom taste profile, then MovieLens recommends other movies for you to watch.



Background Motivation

- Based on the movie watch history of a user(together with ratings), for an unwatched movie, can you predict the ratings that the user possibly gives?
- Recommendation System
  - Recommend potentially high-rating movies
- Multi-information
- Various types of data
- A good fit for our class



UCLA

Task

Based on the information of

- Ratings: UserId, MovieId, Rating, Timestamp
- Tags: UserId, MovieId, Tag, Timestamp
    - Tags are user-generated metadata
- Movies: MovieId, Title, Genres
    - Action, Adventure, Animation etc.
- Links: MovieId, IMDBId, TMDBId
    - Helps you refer to additional information
- Tag Genome: MovieId, TagId, Relevance
    - Relevance indicates how strong a movie is associated with a tag.

We aim to:

Given an unseen UserId-MovieId pair, predict the rating.

UCLA

Task

Training part: Use training data to build your model

- Movie Genre/Tag Genome information
- User tag/rating history ~12M ratings

Validation part: Use validation data to evaluate your model

- user-movie pairs that were unseen in the training part
- ratings for the pairs ~4M ratings

Testing part: Get a score based on the tesing result

- user-movie pairs that were unseen above ~4M tests

UCLA

Dataset

For Our Course Project:

- Based on the MovieLens Latest Datasets
- The corresponding user Ids are re-hashed
  - Do not try to retrieve the original dataset and decode our hashing, you will end up wasting time, trust us :)

UCLA

Submit your results to Kaggle!

Try your model on the Kaggle competition: https://www.kaggle.com/c/movieratepredictions

See your score on the leaderboard

Evaluation:

- RMSE: Root Mean Squared Error
- More details of the Kaggle competition will be released soon

UCLA

Project Grading

- Midterm Report (5% of total)
- Final Report (20% of total)
- Performance on Kaggle
  - 50% of your final report score will be influenced by the performance of your project on the leaderboard(Which is 10% of total).

Project Midterm Report

- **Approximately 3 pages**
- Current progress about project, including
  - Data processing and transformation
  - Designed & tested models / methods
- Discussion and future project plan
  - Some conclusions and findings
  - Analysis of current models and techniques
  - Timeline of future project plan (around 4 weeks)

Details about midterm report guidelines will be released later!

Project Final Report

- **No longer than 10 pages**
- Must include:
  - Group member information
  - Data selection and pre-processing
  - Model and techniques
  - Evaluation and conclusion
  - *Current leaderboard rank and score
  - References and credit (papers, other's codes, maximum 1 page)
  - Related work (maximum ½ page)
  - Task distribution form and peer evaluation form
- ACM paper format
- Must NOT include:
  - Background or too much description on given original datasets
  - Any source code

Details about final report guidelines will be released later!

UCLA

Task Distribution Form: Example

Task	People
Data processing	Student A
Implementation: Algorithm 1	Student B, C
Implementation: Algorithm 2	Student B, D
Implementation: Algorithm 3	Student A, D
Writing final report	Student C

Peer Evaluation Form: Example

	NAMES		
CRITERIA	John	Alice	Bob
Attendance at group meetings	4	4	3
Availability when needed	5	4	3
Highly contributed to writing and proof reading of the final report.	5	5	1
Reliability	5	5	2
Contributed ideas that were of high quality.	4	5	2
Approximately, the amount of time spent on this project was comparable to other group members.	5	5	2
Overall (Would you work with them again?)	5	5	2

Question:
Do you think some member in your group should be given a lower score than the group score? If yes, please list the name, and explain why.

Important Dates & Milestones

- **Jan. 18**: Group formation due (1% in total as participation)
- **Feb. 18**: Midterm project report due
- **Mar. 20**: Final project report due (together with all codes)

UCLA

Q & A

UCLA

Thank you!

Enjoy "mining" and good luck!