



CS6120 Project - Movie Recommendation

TITLE:

GROUP 24: Xuan Zhang, Yi Chen, Meishan Li, Jia Xu, Qia Lin

DATE: 07-03-2022



Introduction

In recent years more and more people are choosing to watch movies on the Internet. The variety and number of movies are also increasing. People often don't know which one to watch instead of facing so many movies. If there is a good recommendation application for movies, these recommendations will help people to choose a movie of the type they want to watch. We will implement a movie recommendation tool by using unsupervised learning models to group movies, and using cosine similarity to select movies the users want to watch more effectively.



Datasets from Kaggle

1. TMDB 5000 Movies

Contains 4083 movies information,
include budget, genres, keywords,
overview, popularity and vote_average.

2. TMDB_5000_credits

Contains movie_id, title, cast, crew



Method, Models

1. Observe Data

- ✓ Tsne

2. Select Features

- ✓ Based on results

3. Vectorize Tokens

- ✓ BoW
- ✓ TF-IDF

4. Group Data

- ✓ K-mean
- ✓ LSH

5. Calculate Cosine Similarity

- ✓ Btwn feature vectors

6. Give recommendations

- ✓ For highest similarity

7. Evaluate model

- ✓ Precision
- ✓ Recall
- ✓ Accuracy
- ✓ RMES

A black clapperboard with a white and black chevron pattern on the top bar. The word "Tools" is written in large white letters on the left side of the board.

Tools

- Sklearn
 - Clustering
 - Preprocessing
- Nltk, Re, String
 - Tokenize
 - Delete stopwords
- Pandas, Seaborn, Matplotlib
 - Observe data
 - Plot pictures about T-sne, RMES, precision, recall and accuracy etc.
- Math
 - Calculate cosine similarity



Milestones

Week 9

- Collect the dataset.
- Clean and preprocess the dataset.
- Observe the data using T-sne etc.
- Select the appropriate features of the data.

Week 10

- Vectorize the tokens from the data.
- Implement BOW method and TF-IDF method

Week 11

- Group data using clustering algorithms (like k-mean) or LSH.

Week 12

- Calculate cosine similarity.
- Generate recommendations.
- Evaluate the model by precision, RMES, recall and accuracy.

Week 13

- Write final report.
- Prepare for the presentation.



Responsi- bilities

Xuan Zhang:

Code, Proposal, Data Preprocessing, etc.

Yi Chen:

Code, Presentation, Vectorize Tokens, etc.

Meishan Li:

Code, Slides, Data Clustering, etc.

Jia Xu:

Code, Report, Calculate Cosine Similarity, etc.

Qia Lin:

Code, Model comparison, Evaluation, etc.
