SmartRecs

Personalized Media Recommendations System Using Collaborative Filtering

CSEN 240 – Machine Learning

Wineel Wilson Dasari

Problem Statement

- Too many movies, too little time.
- Users want personalized recommendations instead of browsing endlessly.
- IMDb has a huge dataset, but no free personalized recommender.
- **Goal:** Recommend relevant movies based on user preferences using ML.

Dataset Overview

- Source: https://datasets.imdbws.com
- Files used:
  - $title.basics.tsv$
  - $title.ratings.tsv$
- Filters applied:
  - Movies only
  - Released after 1980
  - At least 1,000 votes
  - Non-adult content

Development Environment (Why Kaggle?)

- Challenges with Other Platforms
  - **Local PC:** Crashed during processing due to limited RAM
  - **Google Colab:** Free RAM is not sufficient for large matrices
- Why Kaggle?
  - Offers up to **30 GB** of **RAM** in free notebooks
  - Easy to **upload and access large datasets**
  - Smooth execution for **TF-IDF, cosine similarity, and SVD**
- Dataset Size
  - $title.basics.tsv$: ~1.01 GB
  - $title.ratings.tsv$: ~27 MB
  - Total in memory: **Up to 25 GB**

Methodology Overview

- Hybrid Recommendation Approach:
  - Genre-based content similarity (TF-IDF + Cosine)
  - Collaborative filtering using SVD
- Combine both for better relevance

Genre-Based Similarity (Content-Based)

- TF-IDF applied to movie genres (e.g., "Comedy Action Romance")
- Cosine similarity to find movies with similar genre vectors

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity

# Convert genre strings to a space-separated format for TF-IDF input
# Example: "Action,Comedy,Drama" → "Action Comedy Drama"
movies_filtered['genre_str'] = movies_filtered['genres'].fillna('').apply(lambda x: x.replace(',', ' '))

# Create a TF-IDF matrix based on the genre strings
tfidf = TfidfVectorizer()
genre_matrix = tfidf.fit_transform(movies_filtered['genre_str'])

# Compute pairwise cosine similarity between movies based on genre vectors
genre_similarity = cosine_similarity(genre_matrix)
```

Collaborative Filtering with SVD

- Fake users: Movie IDs ($tconst$)
- Fake items: Movie titles
- Ratings: IMDb average rating
- Trained an SVD model to predict unseen movie ratings

Hybrid Score Calculation

- Combine SVD score and genre similarity:
  - $Final\ Score\ =\ 60\%\ SVD\ +\ 40\%\ Genre\ Similarity$

```
final_score = 0.6 * svd_score + 0.4 * genre_score
hybrid_scores.append((row['primaryTitle'], final_score))
```

- Filter out Horror unless explicitly selected
- Smart fuzzy matching for flexible input

Demo Snapshots / Outputs

```
Enter a movie name:  Inception

Did you mean one of these?
  0: Inception (2010.0, Action,Adventure,Sci-Fi)

Enter the number of the correct movie:  0

Showing recommendations based on: Inception
```

```
Top Recommendations:

Title                                                    Year   Genres                                        Score
---------------------------------------------------------------------------------------------------------------------
Edge of Tomorrow                                         2014.0 Action,Adventure,Sci-Fi                       4.45
Vikram                                                   1986.0 Action,Adventure,Sci-Fi                       4.45
Terminator Salvation                                     2009.0 Action,Adventure,Sci-Fi                       4.43
Predators                                                2010.0 Action,Adventure,Sci-Fi                       4.42
Rogue One: A Star Wars Story                             2016.0 Action,Adventure,Sci-Fi                       4.42
The Matrix                                               1999.0 Action,Sci-Fi                                 4.4
Independence Day                                         1996.0 Action,Adventure,Sci-Fi                       4.4
Avengers: Endgame                                        2019.0 Action,Adventure,Sci-Fi                       4.39
Black Panther                                            2018.0 Action,Adventure,Sci-Fi                       4.38
Mad Max Beyond Thunderdome                               1985.0 Action,Adventure,Sci-Fi                       4.38

Recommendation complete.
```

```
Enter a movie name:  cars

Did you mean one of these?
  0: Used Cars (1980.0, Comedy)
  1: Riding in Cars with Boys (2001.0, Biography,Comedy,Drama)
  2: Old Men in New Cars (2002.0, Action,Comedy,Crime)
  3: Cars (2006.0, Adventure,Animation,Comedy)
  4: Serbian Scars (2009.0, Action,Drama,Thriller)
  5: Battle Scars (2020.0, Crime,Drama,War)
  6: Cars 2 (2011.0, Adventure,Animation,Comedy)
  7: Cars of the Revolution (2008.0, Drama,History)
  8: Stealing Cars (2015.0, Drama)
  9: Cars 3 (2017.0, Adventure,Animation,Comedy)

Enter the number of the correct movie:  6

Showing recommendations based on: Cars 2


Top Recommendations:

Title                                                  Year   Genres                                     Score
--------------------------------------------------------------------------------------------------------------
Finding Nemo                                           2003.0 Adventure,Animation,Comedy                 4.29
Isle of Dogs                                           2018.0 Adventure,Animation,Comedy                 4.28
9                                                      2002.0 Crime,Drama,Mystery                        4.27
The Lion King                                          1994.0 Adventure,Animation,Drama                  4.27
The Little Prince                                      2015.0 Adventure,Animation,Comedy                 4.27
Sinbad: Legend of the Seven Seas                       2003.0 Adventure,Animation,Comedy                 4.27
That Christmas                                         2024.0 Adventure,Animation,Comedy                 4.26
Home                                                   2009.0 Documentary,Family                         4.25
The Iron Giant                                         1999.0 Action,Adventure,Animation                 4.25
Ice Age: The Meltdown                                  2006.0 Adventure,Animation,Comedy                 4.24

Recommendation complete.
```

Challenges & Learnings

- IMDb data requires cleaning and deduplication
- Mapping between content and collaborative filters
- Handling input in Kaggle (input prompt issue)
- Learned: TF-IDF, SVD, hybrid logic, fuzzy matching

Conclusion & Future Work

- Accurate and genre-safe recommendations
- Future improvements:
  - Include actors/directors in similarity
  - User login for long-term personalization
  - Web app interface

Thank you…