Hybrid Movie Recommendation System

Introduction

The Hybrid Movie Recommendation System project aims to develop a robust recommendation engine that leverages both user-item interactions and item attributes to suggest movies tailored to individual user preferences. By combining collaborative filtering and content-based filtering techniques, the system provides more accurate and personalized recommendations.

Objectives

- Preprocess and clean the MovieLens 100K dataset for analysis.
- Implement content-based filtering using movie metadata.
- Implement collaborative filtering using matrix factorization.
- Integrate both filtering techniques into a hybrid model.
- Create an interactive user interface for movie recommendations.
- Evaluate the system's performance using standard metrics.

Data Ingestion and Preprocessing

The MovieLens 100K dataset was loaded and preprocessed to ensure data consistency and handle missing values. The dataset includes user ratings, movie details (e.g., genres), and user information. Key steps included:

- Loading movie, rating, and user data from the dataset.
- Merging ratings with movie titles and genres.
- Handling missing values and ensuring ratings are within the valid range (1 to 5).
- Visualizing the distribution of ratings to understand user behavior.

Content-Based Filtering Module

This module focuses on recommending movies based on their attributes, such as genres and titles.

- **Feature Extraction**: Movie genres were combined with titles to create a text feature for each movie.
- **Similarity Computation**: TF-IDF vectorization was applied to the text features, and cosine similarity was computed to find similar movies.
- **Recommendation Generation**: For a given movie, the system recommends the top N most similar movies based on cosine similarity scores.

Collaborative Filtering Module

This module predicts user preferences based on historical user-item interactions.

- **Matrix Factorization**: The Surprise library's SVD algorithm was used to factorize the user-item rating matrix.
- **Rating Prediction**: The trained SVD model predicts ratings for unseen movies, enabling the system to recommend movies likely to be rated highly by the user.

Hybrid Recommendation Engine

The hybrid model combines the strengths of both content-based and collaborative filtering.

- **Integration Strategy**: Recommendations from both models are combined using a weighted averaging approach.
- **Weight Adjustment**: Weights are adjusted based on validation performance to optimize the hybrid model's recommendations.

User Interface

An interactive interface was developed using Streamlit, allowing users to:

- Input their user ID or select movies they like.
- Receive personalized movie recommendations based on the hybrid model.
- View movie posters and details for a more engaging experience.

Evaluation Metrics

The system's performance was assessed using several metrics:

- **Root Mean Square Error (RMSE)**: Measures the accuracy of predicted ratings.
- **Mean Absolute Error (MAE)**: Provides another perspective on prediction accuracy.
- **Precision, Recall, and F1-Score**: Evaluates the relevance of recommendations, focusing on top-N recommendations.

The hybrid model demonstrated improved performance over individual filtering techniques, as evidenced by the evaluation metrics.

Conclusion

The Hybrid Movie Recommendation System successfully integrates content-based and collaborative filtering to provide personalized movie suggestions. Future enhancements could include incorporating additional features (e.g., movie overviews, user demographics) and exploring advanced hybrid techniques (e.g., deep learning-based models).