

Netflix Movie Recommendation System

A Machine Learning Case Study

Rohit Karwa

1 Problem

Netflix aims to connect users with movies they love by providing accurate movie recommendations. Their in-house system, *Cinematch*, predicts user ratings based on previous interactions. However, Netflix is exploring alternate machine learning-based approaches that can improve upon Cinematch's predictions by at least 10% in terms of accuracy, measured via metrics like RMSE. The challenge involves leveraging collaborative and content-based filtering techniques to build a better recommendation engine.

2 Approach

The following steps were followed in the project:

1. Data Exploration and Preprocessing:

- Parsed rating data with 480,189 users and 17,770 movies.
- Ratings are integers from 1 to 5, along with timestamp information.

2. Modeling Techniques:

- **Collaborative Filtering:** Using matrix factorization with Singular Value Decomposition (SVD), SVD++, and KNN-based models from the `surprise` library.
- **Content-Based Filtering:** Feature-based recommendations using movie metadata (if available).
- **Hybrid Models:** Ensemble of collaborative and content-based predictions using XGBoost regression.

3. Evaluation:

- Used RMSE and MAPE to measure prediction accuracy.
- Benchmarked models against the baseline Cinematch system.

3 Tech Stack

- **Language:** Python 3
- **Libraries:**
 - `surprise` - For collaborative filtering models
 - `XGBoost` - For regression and ensemble learning
 - `scikit-learn` - Machine learning utilities
 - `pandas`, `numpy`, `scipy` - Data handling and numerical operations
 - `seaborn`, `matplotlib` - Visualization
- **IDE:** Jupyter Notebook (via Anaconda)

4 Results

- Achieved a significant reduction in RMSE compared to the baseline.
- SVD++ and hybrid models outperformed basic algorithms.
- Recommendations generated are personalized, scalable, and interpretable to some extent.
- Model results align with insights from Netflix's own research (Koren et al.).

Acknowledgments: Applied AI Course, Netflix Prize data, Netflix Prize, and Surprise Library.