

Multimodal Movie Recommendation System - Progress Update Report - yz10077 - 3033-061

(A) Technical Progress Summary

1. Data Preprocessing

My project focuses on developing a comprehensive multimodal movie recommendation system using the MovieLens 100k dataset as the foundation. The data preprocessing phase involved several key components:

- **Dataset Integration:** Successfully integrated MovieLens 100k ratings data (100,000 ratings, 943 users, 1,682 movies) with additional metadata from TMDB and IMDb APIs
- **Feature Engineering:** Implemented comprehensive user feature extraction including age groups, gender, occupation, and geographic regions, resulting in 51-dimensional user features after preprocessing
- **Multimodal Data Collection:** Downloaded movie poster images to `multimodal_images/` folder and trailer videos to `Trailers/` folder for visual feature extraction
- **Dimensionality Reduction:** Applied both PCA and SVD techniques for feature optimization, reducing user features to 18 dimensions and movie features to 13-18 dimensions while maintaining 85-90% explained variance

2. Modeling

I implemented and compared 14 different recommendation algorithms across three main categories:

- **Collaborative Filtering Models:** Developed both user-based and item-based collaborative filtering approaches with optional feature integration
- **Feature-Enhanced Models:** Created variants incorporating PCA and SVD-reduced user/movie features
- **Hybrid Systems:** Combined multiple approaches with weighted fusion strategies
- **Multimodal Integration:** Implemented vision transformer (ViT) for image feature extraction and BERT for text processing, though this component showed challenges in current results

3. Experimental Results

The comprehensive evaluation revealed several key findings:

- **Best Performance:** Hybrid recommendation system combining SVD-enhanced user features with rating-only item features achieved the lowest RMSE of 0.8892
- **Feature Value Analysis:** User features (both PCA and SVD) provided modest improvements (0.64-1.67% RMSE reduction)
- **Algorithm Comparison:** Hybrid models consistently outperformed single-approach methods, with performance improvements of up to 10.16% over baseline user-based collaborative filtering
- **Multimodal Challenges:** Current multimodal model performance (RMSE/MAE) is suboptimal due to dimensionality mismatch between limited movie count and high-dimensional visual/text features, leading to popularity bias where highly-rated movies dominate recommendations

(B) TA Meeting Discussion Summary

On July 31st at 10:20 AM, I met with TA Rasmika Billa via Google Meet to discuss project progress for my predictive analysis course. During the meeting, I presented the current workflow and demonstrated the multimodal data collection efforts, including the downloaded movie trailers in the `Trailers/` folder and poster images in the `multimodal_images/` folder. I reviewed the comprehensive experimental results documented in the Jupyter notebooks, discussing the data preprocessing pipeline, feature selection strategies, and model building approaches.

I explained the current limitation where multimodal models are not performing as expected due to the relatively small number of movies (1,682) compared to the high dimensionality of visual and textual features, which creates a training challenge where highly-rated popular movies are consistently recommended regardless of the additional metadata. I discussed potential future improvements including data augmentation strategies, feature selection optimization, and alternative fusion mechanisms to better leverage the multimodal information while addressing the dimensionality imbalance issue.