Task 1: Guided Active Learning Setup for Cold-Start Recommendation

June 23, 2025

1 Introduction

This document outlines the guided active learning setup developed for addressing the cold-start problem in movie recommender systems using the MovieLens dataset. The method involves soliciting user ratings while providing feedback on how these ratings influence future recommendations.

2 Method Description

The system employs an interactive approach where new users rate movies from a curated list of 10 popular movies, selected based on a minimum of 50 ratings and an average rating above 3.0. Users provide scores from 0.5 to 5, reflecting either their past rating or interest level.

2.1 Active Learning Mechanism

The initial movie set is diversified by ensuring minimal genre overlap, computed by maintaining a set of unique genres and selecting movies where the intersection with existing genres is less than 2. This ensures variety and broad taste coverage.

2.2 Guided Feedback

For the experimental group, each movie includes an explanation, e.g., "This [genres] movie helps us understand your taste in [primary genre] films." This feedback aims to visualize the impact of ratings, encouraging strategic input. The control group receives no such explanations, serving as a baseline.

2.3 Recommendation Algorithm

The recommendation leverages Non-Negative Matrix Factorization (NMF) on the user-item matrix. The optimization problem is defined as:

$$\min_{U,V} \sum_{(i,j)\in\mathcal{K}} \left(R_{ij} - U_i^T V_j\right)^2 + \lambda \left(\|U\|_F^2 + \|V\|_F^2\right)$$

where R_{ij} is the rating, $U_i \in \mathbb{R}^K$ and $V_j \in \mathbb{R}^K$ are latent representations, and $\lambda = 0.01$ is the regularization parameter. For a new user, U_i is solved via:

$$\min_{U_i} \sum_{j \in \mathcal{K}_i} \left(R_{ij} - U_i^T V_j\right)^2 + \lambda \|U_i\|_F^2$$

using 20 latent components, initialized randomly with a fixed seed for reproducibility.

3 Implementation Details

The system loads the MovieLens small dataset (100k ratings) and processes it using pandas and scikit-learn. The interface dynamically updates with new movie sets via the ‘next$_q uestions‘ view, ensuringcor$