Course: COMP9727 Recommender Systems
Proposer: Zihao Zhou
zID: z5423698
Date: July 4, 2025

Topic: Anime Recommender System

1. Project Core Idea & Aims

This project hopes to design and implement an intelligent anime recommendation system for a large community of anime fans. Unlike mainstream movies or music, the appeal of anime is often driven by very specific attributes, such as a particular genre, art style, production studio, or voice actor. Recommender systems based only on general ratings often fail to capture these nuanced preferences.

The core goal of this project is to build a hybrid recommendation model. It will not only utilize user rating data, but also deeply integrate the rich metadata of the anime works themselves. The project hopes to provide accurate recommendations and help users discover niche and quality works. This system can effectively help users when they have common problems such as “too many new dramas and don't know what to watch” and “depression after watching anime” after watching their favorite works.

2. Scope

- **Domain:** Anime and Fan Communities
- Users: Our target users are participants in the anime community who love anime and enjoy exploring different genres.
- Content: The system will recommend anime works (including TV series, movies, OVAs, etc.) that users may be interested in but have not yet watched or rated.
- **Interface Prototyping:** Although we will not be developing a full front-end, the design concept for the system is centered around the user's personal home page. This page will contain a “Recommended for you” module, which will display a list of recommendations in the form of cards with anime covers. Users will be able to provide **clear feedback** by clicking through to the details page or by giving a 1-10 star rating directly.
- **Cold Launch Strategy**:
  - **User Cold Launch**: for new users, the system can guide them to rate approximately 10 well-known anime titles (e.g. Attack on Titan, Ghostbusters, Alchemist of Steel). Using this initial data, the **content-based recommendation** method can immediately generate the first personalized recommendations.
  - **Commodity Cold Launch**: when a new anime series starts airing, it does not yet have user ratings. Our system will utilize its known metadata, such as **Type, Genre,**

and Studio, to recommend it to users who have had a preference for these attributes in the past, thus effectively solving the “new merchandise” problem.

3. Dataset & Exploratory Data Analysis

- **Selected Dataset: Anime Recommendations Database (from MyAnimeList.net)**
  - **Source:** Kaggle (https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database)
  - **Reasons for selection**
    1. **Rich metadata**: The dataset is extremely rich, containing not only ratings, but also high-quality structured data, such as “Genre”, ‘Episode’, “Rating”. . All of these are very suitable for implementing a hybrid recommender system. Among them, “rating” is very suitable for realizing hybrid recommender system.
    2. **Community Driven**: This data comes from one of the largest anime fan communities in the world, which ensures that users' ratings and behaviors truly reflect their preferences.
- **Exploratory Data Analysis (EDA) - Preliminary Plan**: before we start modeling, we will explore the dataset so as to obtain key information:
  1. **RATING DISTRIBUTION**: Plot a histogram of 1-10 star ratings. We predict that users will, with high probability, give high ratings to their favorite works, as well as very low ratings to works they hate.
  2. **Genre Analysis**: Analyze and visualize some of the most popular anime genres (e.g., action, comedy, sci-fi) to visualize the preferences of mainstream viewers.
  3. **Genre Analysis**: Compare the number of titles and average ratings of different genres, such as TV series, movies, and OVAs.
  4. **User Rating Behavior**: analyze the rating distribution of each user to identify average and core users, which will help to understand the data more deeply.

4. Methodology

We will adopt a phased, iterative approach to build our system, ensuring that each step has a solid baseline for comparison.

- **Phase 1: Baseline Models**
  - **Objective:** Establish the simplest models to measure the performance lift of all subsequent, more complex algorithms.
  - **Methods:**
    1. **Global Top-N:** Recommend the N anime with the highest average rating across the community to all users.
    2. **Simple Content-Based:** Use only the anime's **genre** features. By calculating the Jaccard similarity of genres, we can recommend items that are most similar to a user's high-rated anime.
- **Phase II: Collaborative Filtering**

- **Objective**: to implement collaborative filtering algorithms, which is the core of personalized recommendation.
- **Methodology**: We will implement Item-Based CF on the data of user ratings of anime. We chose this method because the number of anime is relatively stable and the similarity between items can be pre-calculated offline, thus improving efficiency.

- **Phase III: Matrix Decomposition**
  - **Objective**: improve prediction accuracy by utilizing a more robust latent factor model.
  - **Methodology**: implement a recommendation algorithm based on **SVD (Singular Value Decomposition)**. The model learns low-dimensional latent features of users and anime to achieve more accurate rating predictions.
- **Phase 4: Hybrid Recommendation**
  - **Objective**: this is the core innovation of our project. We will combine the personalization feature of collaborative filtering with content-based features to build a more powerful model and solve the cold-start problem.
  - **Methodology**: we plan to adopt the **feature combination** strategy. Specifically, we will use a library similar to **LightFM**, which can incorporate item metadata (e.g., **Type, Genre, Studio**, etc.) as auxiliary information during the matrix decomposition process into the training process. This allows the model to not only learn from ratings, but also understand concepts such as “a user who likes Studio A's sci-fi anime may also like Studio B's similar work”, thus significantly improving the depth and breadth of recommendations.

5. Evaluation Plan

We will conduct a comprehensive and rigorous evaluation of our model, covering both offline metrics and user studies.

- **Offline Evaluation:**
  - **Methodology:** The scoring dataset will be divided into training and test sets in a ratio of 80/20.
  - **Evaluation metrics:**
    1. **Prediction Accuracy:** We will use **Root Mean Square Error (RMSE)** to evaluate the accuracy of our model in predicting user ratings.
    2. **Ranking Accuracy:** We will use **NDCG@10** to evaluate the quality of the Top-10 recommendation list.NDCG is a very effective metric because it rewards models that place more relevant items higher on the list.
    3. **Novelty and Coverage:** We will also measure the **Novelty** (the ability to recommend non-obvious items) and **Catalog Coverage** of the recommended listings to ensure that our model avoids the “popularity bias” trap.
- **User Research Design (Part III):**

- **Objective:** Evaluate users' subjective perceptions of the different models' recommendations, especially their satisfaction with our final hybrid model.
- **Design (A/B testing):**
  1. 10-15 anime fans were recruited as participants. 2.
  2. Randomly divide them into two groups: Group A will use recommendations generated by **standard collaborative filtering** models, while Group B will use recommendations generated by our **hybrid models**. 3.
  3. present each participant with their generated Top 10 list and ask them to complete a short questionnaire.

- **Questionnaire items (Likert Scale, 1-5):**
  - "How well does this list match your preferences?" (Accuracy)
  - "How many works on this list have you not seen before but are now interested in?" (Novelty/unexpected findings)
  - "Do you feel that this list covers a wide range of styles that you might like?" (Diversity)
  - "Overall, how satisfied were you with this recommendation experience?" (Overall satisfaction)
- **ANALYSIS**: By comparing the scores of the two groups, we can quantitatively demonstrate whether our hybrid model provides a better user experience than the traditional model, which will be a strong point of analysis in the individual reports.