Project Design

Zid: z5502296

Name: XIAOYU BAI

1. Scope

1.1 Project Background and Motivation

This is a movie recommendation system designed for online platform users who enjoy watching movies. It is mainly deployed on online movie platforms to address the problem of users' difficulty in making choices under information overload.

The movie content is rich and updated frequently. Platform users often face the dilemma of "wanting to watch a movie but not knowing what to watch". Therefore, there is an urgent need for an intelligent recommendation tool to shorten the decision-making time, discover potential interests, and enhance the satisfaction and stickiness of users in using the platform.

This system focuses on the field of movie recommendation and targets users including movie enthusiasts, ordinary viewers who frequently use movie websites, and new registered users. It is particularly suitable for young people, office workers, students, and other groups with high movie-watching frequency but limited time. By integrating collaborative filtering and content recommendation methods, this system can provide more balanced and intelligent solutions in areas such as cold start, interest drift, and diversity recommendation.

Therefore, this project aims to design and implement a hybrid recommendation system that integrates strategies such as collaborative filtering (CF) and content-based recommendation (CB), to enhance the effectiveness of personalized recommendations. The system also considers both user experience and the feasibility of system engineering implementation and possesses strong practical application capabilities.

1.2 Application scenario

This system is a hybrid recommendation system for movie websites, designed for registered users who wish to receive personalized viewing suggestions. Users browse the recommendation cards on the platform and can provide feedback through interactive methods such as "like/dislike". The recommendation process is triggered when users log in or refresh the page. The overall interaction process is simple and intuitive.

The recommendation system is mainly deployed on movie streaming platforms, targeting registered users of the platform, especially daily active users (DAU). Through precise recommendations, it aims to increase the average user stay time, click-through rate, and conversion rate.

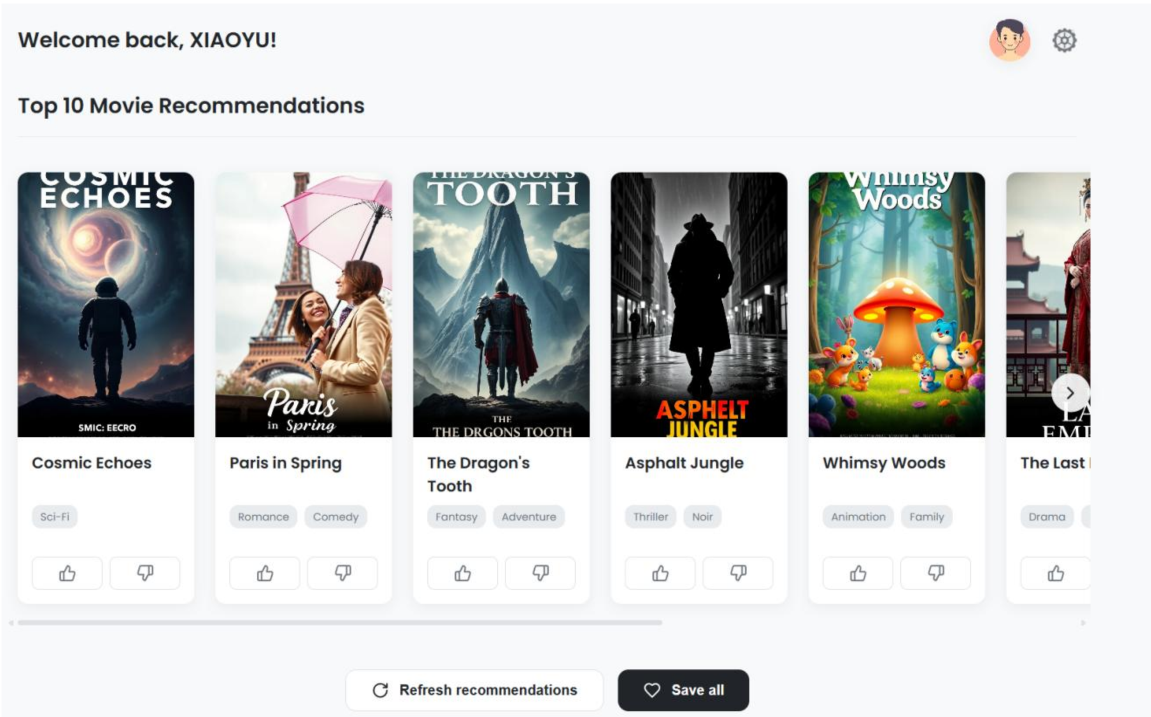
When users first log into the system, they will receive an initialization questionnaire recommendation (mainly in the form of CB). Subsequently, based on user ratings and behavioral data, a profile is constructed, and the recommendation ranking is dynamically adjusted. The system should support cold start for new users, cooling of popular content, and optimization of diversity for long-tail content. The recommendation system is mainly deployed on movie streaming platforms, targeting registered users of the platform, especially daily active users (DAU). Through precise recommendations, it aims to increase the average user stay time, click-through rate, and conversion rate

When a user logs into the system for the first time, they will be recommended an initialization questionnaire (with CB as the main method). Subsequently, based on the user's ratings and behavioral data, a profile will be constructed, and the recommendation ranking will be dynamically adjusted. The system should support cold start for new users, cooling of popular content, and optimization of diversity for long-tail content.

1.3 Recommended interface design

The user interface of this recommendation system is mainly presented to users through web pages. The platform is a module of an online movie website. Users can access the recommendation system interface by visiting the homepage or the "Recommended for You" section. The overall UI design is simple and modern, following the card-style content display logic of Netflix and Spotify, to ensure usability and visual appeal. The recommendation interface adopts a card layout, presenting a total of 10 recommended movies. Each recommended card includes:

- Cover poster (click to enter the detail page)
- Movie title and rating label
- Like / Dislike Two feedback buttons



The top part of the interface features a welcome message and a user avatar area, while the bottom section contains the buttons " Refresh Recommendations" and " Save All".

1.4 User interaction and feedback mechanism design

This recommendation system establishes connections with users through various interaction methods and collects feedback data to improve the recommendation effect. Each piece of recommended content in the system interface is accompanied by an explicit feedback button, such as "Like", "Dislike", "Save", "Refresh", and simultaneously implicitly collects users' browsing, clicking and stay time behavior characteristics.

The detailed feedback explanation is as follows：

- Like: Strengthen the weight of the characteristics of the movie in the user portrait, and improve the probability of recommending similar content;
- Dislike: hide the movie in the current session and reduce the probability of recommending similar content;
- Collection: regarded as a strong preference signal, used to directly train user interest embedding;
- Refresh recommendations: the system records the user's skipping behavior and feeds back the non-relevance signal of the current recommendation ranking;

Based on the above interaction feedback, the system performs a batch model update daily to optimize user interest vectors, adjust similarity calculation logic, and update the SVD latent factor matrix. In the future, it can expand into a near real-time online learning architecture, allowing for rapid fine-tuning of model parameters after interactions. The system collects feedback through explicit (likes/dislikes) and implicit (browsing, clicking, and dwell time) methods. User behaviors are collected via front-end event listeners, uploaded to the Kafka message queue, and then asynchronously consumed and recorded in the user behavior log by the backend.

The recommendation model is updated once a day, including user interest vector reconstruction, cold start user portrait generation, and model weight fine-tuning. The update method adopts batch processing offline to ensure production stability.

1.5 Business model and potential for implementation

The recommendation system has good business expansion space, and its business logic is not limited to the technology laboratory but has the imagination and application value of realization.

As the recommendation engine in the movie website, this recommendation system can directly support the following business models:

Advertising realization (CPC model): The platform cooperates with online movie ticketing or peripheral content platforms. After users click on the recommendation items, they will be redirected to the ticket purchase/detail page. The platform can charge advertising fees according to the number of clicks to realize the realization of recommended traffic.

Premium membership service (subscription system): only ordinary recommendation is provided to non-members, and advanced functions such as "personalized recommendation" and "star special recommendation" are opened for member users to promote user payment subscription. The more accurate the recommendation effect is, the higher the conversion rate of members will be.

Cold start film promotion entrance (exposure diversion): As new independent films or documentaries are difficult to rank due to lack of historical ratings, the platform can guide the film producers to pay for recommended exposure positions and assist them to accurately reach the target audience group through cold start recommendation algorithm.

Brand cross-border linkage recommendation: for example, when users watch food documentaries, they can be linked to recommend discount coupons of takeout platforms, or after watching comedy movies, they can be recommended tickets for comedy performances and e-commerce joint products, so as to open up the upstream and downstream business links.

2. Datasets

2.1 Data set selection and processing process

The project selected the public data set "The Movies Dataset" from the Kaggle platform, which was derived from the TMDB data interface and had a large scale and real distribution.

Core data tables:

movies_metadata.csv: The main Movies Metadata file. Contains information on 45,000 movies featured in the Full MovieLens dataset. Features include posters, backdrops, budget, revenue, release dates, languages, production countries and companies.

keywords.csv: Contains the movie plot keywords for our MovieLens movies. Available in the form of a stringified JSON Object.

credits.csv: Consists of Cast and Crew Information for all our movies. Available in the form of a stringified JSON Object.

links.csv: The file that contains the TMDB and IMDB IDs of all the movies featured in the Full MovieLens dataset.

links_small.csv: Contains the TMDB and IMDB IDs of a small subset of 9,000 movies of the Full Dataset.

ratings_small.csv: The subset of 100,000 ratings from 700 users on 9,000 movies.

The Full MovieLens Dataset consisting of 26 million ratings and 750,000 tag applications from 270,000 users on all the 45,000 movies in this dataset can be accessed here

2.2 Feature selection and preprocessing

- The scoring matrix is sparsified into User-Item Matrix for CF model;
- JSON parsing and normalization processing are carried out for genres, keywords, cast and other fields;

- Clean, segment and TF-IDF encode the text fields (such as introduction) for content modeling;
- De-noise the rating records (such as filtering extreme user behavior).

2.3 Data quality assessment

The "The Movies Dataset" data set used in this project is extremely abundant in terms of data scale and dimension, and is suitable for training collaborative filtering and hybrid recommendation models. This dataset contains approximately 26 million rating records, involving about 270,000 users and 45,000 movies. Compared with the small sample data used in the course tutorials, this dataset provides a more comprehensive user-item interaction matrix, enabling effective training of latent factor models and providing a good foundation for content modeling and hybrid recommendation.

2.4 Data Limitation Statement

Despite the rich fields and large number of interactions in The Movies Dataset, there are still the following limitations:

- Non-complete data: This data set is from Kaggle and only contains part of the active user data of the TMDB platform. It lacks complete user distribution and long-tail interaction behavior, so it cannot fully restore the cold user problem in the real world.
- The data has been pre-cleaned: Kaggle data has eliminated abnormal items and missing fields, reducing the complexity of the data. In the actual deployment environment, the recommendation system needs to deal with noise scores, malicious behaviors and other problems, which are not reflected in this data set.
- The task setting is biased towards rating prediction: the structure of this data set is more suitable for rating prediction than for production-level tasks such as complete ranking, click-through rate optimization, and real-time recommendation.

3. Methods

The recommendation system adopts multi-model fusion architecture, combining collaborative filtering and content-based recommendations to improve the overall robustness and coverage.

3.1 collaborative filtering （CF）

Two approaches are used:

- Item-based CF: Generate a matrix of similar movies by calculating the similarity of ratings between movies;
- SVD++ (Hidden Semantic Model): A low-rank matrix decomposition model is constructed based on Surprise toolkit to effectively alleviate the sparsity problem;

The module performs well when the old user's rate well but is ineffective for cold start users.

3.2 Content-based recommendation (CB)

The attributes of the movie, such as type, key words, language and introduction, are transformed into content vectors to construct user interest portraits.

- TF-IDF + One-hot is used to encode content fields
- The user profile is made up of the average of historical likes
- Cosine similarity is used to calculate interest matching degree

The CB model is user-friendly for new users and long tail content and is suitable for the start-up phase.

3.3 Hybrid recommendations

Integration scoring method:

$$\text{Score} = \alpha \cdot \text{Score\_CF} + (1 - \alpha) \cdot \text{Score\_CB}$$

In the fusion strategy, the α value can be optimized by grid search, cross-validation, and other methods.

Cold start α = 0 (all content recommendation)

Old users α take a medium weight (e.g., 0.5)

4. Evaluation

This chapter builds a complete multi-indicator evaluation framework from the perspective of model layer (Offline Evaluation) and system layer (User-centered Evaluation) and explains the calculation efficiency and experimental design.

4.1 Objectives and Evaluation Dimensions

To comprehensively assess the effectiveness of the recommendation system, two key evaluation dimensions are used:

Assessment Level	Purpose	Key Evaluation Question
Model layer	Evaluate the predictive accuracy of recommendation algorithms using historical interaction data.	How accurately can the model recover users' true preferences and rank relevant items in the Top-N list?
System level	Assess overall user experience and business value, including behavioral feedback and subjective perception.	Are the recommendations accepted and clicked by users? Do they provide enough novelty and diversity?

Use "each user" as the evaluation unit and report the averaged performance across users. The recommendation list size is fixed at Top-10 throughout.

4.2 Model layer metrics

To evaluate the algorithm performance based on historical data, we adopt the following metrics:

Metric	Formula / Description	Sampling Strategy
Precision@10	Proportion of top-10 recommended items that are actually liked by the user. Measures accuracy.	Leave-one-out: reserve last known interaction for test.
Recall@10	Proportion of liked items that appear in the top-10 recommendation. Measures coverage of relevant items.	Same as above.
NDCG@10	Normalized discounted cumulative gain. Rewards relevant items that appear higher in the ranked list.	Same as above.
MAP@10	Mean of Average Precision over all users. Emphasizes correct ranking across multiple relevant items.	Same as above.
Coverage	Proportion of unique movie genres recommended over total genres. Indicates system's ability to go	Aggregated over all users.

Metric	Formula / Description	Sampling Strategy
	beyond popular items.	
Diversity	Average pairwise dissimilarity (e.g., cosine distance using TF-IDF on genres or keywords) among top-N recommended items.	TF-IDF cosine distance computed on textual metadata.

Reserve each user’s last interaction as the test set, and randomly split the remaining 90% as training and 10% as validation.
For cold-start simulation, 5% of new movies are excluded from training and only appear in the test set.

4.3 System-level metrics

To evaluate user-perceived effectiveness of the system, it include both behavioral signals and subjective feedback:

Dimension	Behavioral/Subjective Scale	Improved Indicator Explanation
Click rate (CTR)	Clicks / Impressions	Indicates the attractiveness of the recommendation. Higher CTR means users are interested enough to explore.
Skip rate	Skips / Impressions	Reflects non-relevance or disinterest. A high skip rate may signal mismatch between content and user interests.
Save rate	Saves / Impressions	Represents strong user preference. Saved items may indicate high engagement or intent to revisit.
Novel click ratio	New content clicks / Total clicks	Measures the exploration capacity of the system — whether users are being exposed to new, unseen items.
Satisfaction score (Q1)	Likert scale (1 to 5)	Captures subjective user satisfaction: “How relevant and useful were the recommended items to you?”
Diversity perception	Likert scale (1 to 5)	Captures perceived diversity: “Did the recommendations cover

Dimension	Behavioral/Subjective Scale	Improved Indicator Explanation
(Q2)		a variety of content types?”

Behavioral data (CTR, skip, save) is collected via interaction logging. Subjective data (Q1–Q2) is collected via user questionnaires after exposure to simulated recommendation interfaces.

4.4 Indicator trade-off and main indicator selection

Main index: NDCG@10 (accuracy and ranking)
Secondary indicators: Diversity + Coverage (avoid list convergence)
Strategy: If the improvement of NDCG is less than 1% and the improvement of Diversity is greater than or equal to 10%, it is considered an improvement; otherwise, the original model is maintained.

4.5 Computational efficiency and update strategies

Offline phase: Spark batch retraining SVD++ and updating content vector every morning, which takes less than 60 minutes.
Online phase: Faiss neighbor recall + Redis cache rearrangement; single request P95 delay is less than or equal to 150 ms.

4.6 User Research Experiment Design

Prototype tool: Figma Click-through high-fidelity interface to simulate likes, dislikes, and skips.
Subjects: 10 volunteers were randomly assigned to 3 rounds (Item CF/SVD++/Hybrid) in a random order.
flow path：Show Top 10 → Record click / skip behavior → Complete the questionnaire (Q1~Q5 satisfaction, Q6 model preference).
Data output: Click log + Likert scale; Friedman test is used to compare the differences in satisfaction among the three models.

4.7 Summary

The evaluation framework covers three dimensions: algorithm accuracy, recommendation diversity and real user experience, and considers online computing efficiency, providing a complete quantitative basis for subsequent model iteration and business decisions.