

Project Report

Project Name:

Intelligent Movie Recommendation System

Group 2:

Zeng Zijing A0231548E

Wu Yichen A0231544M

Menu

1 Background	4
1.1 Necessity of the recommendation system	4
1.2 Generation of movie recommendation system	5
2 Research on Existing Recommender System	6
2.1 China's mainstream movie recommendation system	6
2.1.1 Youku's movie recommendation system	6
2.1.2 Douban movie recommendation system	7
2.1.3 LeTV movie recommendation system	8
2.2 International mainstream movie recommendation system	10
2.2.1 Jinni Movie recommendation system	10
2.2.2 IMDB movie recommendation system	11
2.2.3 Rotten Tomatoes movie recommendation system	12
2.2.4 Movielens movie recommendation system	12
3 Analysis of Problems in China's Movie Recommendation System	13
4 Commercial Value of the Project	14
4.1 Project technical advantages	14
4.1.1 Application of tag	14
4.1.2 Application of scoring system	15
4.2 Commercial Value	16
5 System Design	17

5.1 Architecture Design	17
5.2 Database Design	17
5.2.1 User Table	17
5.2.2 Movie Table	18
5.2.3 Ranking Table	18
5.3 Recommendation Design	18
5.3.1 Recommend Process	18
5.3.2 User-based Collaborative Filtering	19
5.3.3 Item-based Collaborative Filtering	20
6 Operating Environment	22
6.1 Introduction of MySQL	22
6.2 Introduction of Django	23
7 User Guide	23
7.1 Install project dependencies	23
7.2 Build Database	24
7.3 Operating System	24

1 Background

1.1 Necessity of the recommendation system

Due to the Internet has the characteristics of open information exchange, bidirectionality of information transmission, and extensive information coverage, the amount of information in the network expands and increases exponentially, and the problem of information overload and information trek on the network is becoming more and more serious.

Solving these problems, the key is to transform the Internet from passively accepting the requests of viewers to actively perceiving the information needs of viewers, and to realize the active information service of the network system to the viewers. In order to provide people with satisfactory information and services, the recommendation technology has emerged as the times require.

Many scholars and Internet users are concerned about the core technology recommendation system by predicting the user's preference for information resources to filter information, according to the specific needs of users through collaborative filtering and other technologies for personalized recommendation. According to this basic principle, the recommendation system has produced a variety of Different algorithms Among many recommendation algorithms, collaborative filtering recommendation is by far the most successful and widely

used personalized recommendation technology. The concept of collaborative filtering is widely used in various fields, such as e-commerce, social network, book recommendation, movie recommendation, etc. Recommendation systems include Amazon.com, Grouplens, and Ringo; domestic recommendation systems have also been in a period of rapid development in recent years, such as Taobao Jingdong Mall.

1.2 Generation of movie recommendation system

At the beginning of the development of video sites, video recommendations came into being. Video recommendation provides users with personalized video recommendations based on current popular videos and users' personalized data, thereby increasing user retention and website traffic. This is one of the most important functions of major video websites. For online movie providers, the recommendation efficiency of the online movie recommendation system will directly affect the company's economic benefits, and at the same time have an important impact on the company's development.

The 2006 Netflix competition was a landmark event in the video recommendation field. The competition offered a reward of 1 million US dollars. It is hoped that researchers can increase the prediction accuracy of Netflix's recommendation algorithm to 10%. Three years after the competition was held, AT&T researchers Take away the prize. This competition attracted many teams to participate and applied collaborative filtering, association rules, singular value

decomposition (SVD) and many other recommendation methods to the field of video recommendation, achieving very good recommendation results. This incident reflects the importance of video websites to the recommendation system. At the same time, companies such as YouTube are also conducting special research in the field of video recommendation, which shows that the recommendation system plays an important role in video websites.

2 Research on Existing Recommender System

2.1 China's mainstream movie recommendation system

2.1.1 Youku's movie recommendation system

We can analyze it through the recommended quality of Youku when we stand at the point of users. The user account is an account registered in 2009. The videos watched and recommended videos are shown in Figure 2-1-1 and Figure 2-1-2.

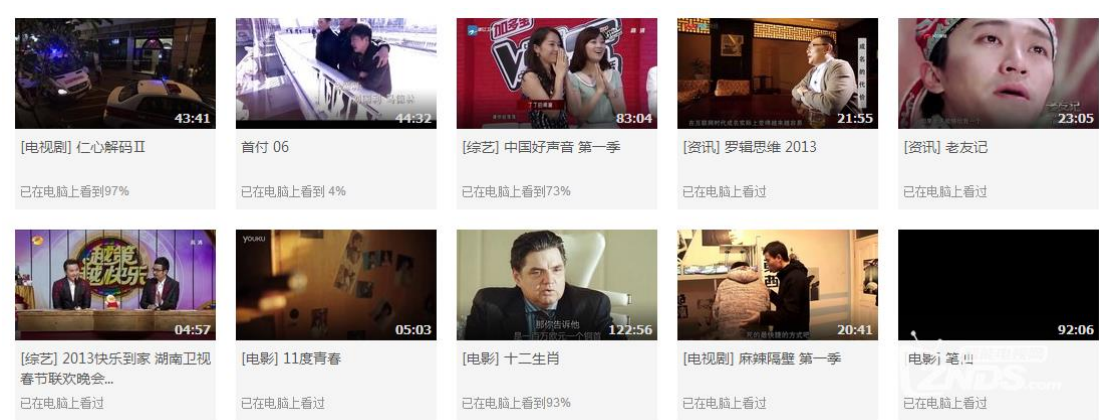


Figure 2-1-1 Watched videos



Figure 2-1-2 Recommended videos

As for the recommendation results, most of the recommendation results have a high degree of similarity with the videos in my viewing records (the viewing records only list the recent viewing records), and this recommendation effect can be achieved through the Content-based method. However, there is a certain degree of overlap with the watched record, and the generalization performance of video recommendation is relatively poor.

2.1.2 Douban movie recommendation system

If the CF algorithm is purely used in movie recommendation, it will produce more serious head effects (popular movies are also recommended). After a user has watched an action movie in a theater, he often rates many movies that are shown in the theater at the same time. If the CF algorithm is directly applied to the user's rating data, it may cause embarrassment for those who like action movies and drama movies.

Douban uses high-quality label data to solve this problem. Firstly, the movies

are grouped according to movie tags (such as action movies, drama movies, etc.); then the CF algorithm is used in each group to generate recommendation results; finally, the recommendations obtained in each group are combined in a weighted combination. It can help people easily describe and classify content for easy retrieval and sharing. Tag has become an important factor in web 2.0. It can easily and effectively define the characteristics of movies on Douban, and can simplify and standardize the characteristics of a certain entity with a few or dozens of tags.

2.1.3 LeTV movie recommendation system

We used a Tencent account to register and associate the account. Among them, the playback record mainly selects the two channels "American TV" and "Sports", and some noise data (the first two videos watched) are added, as shown in the figure2-1-3.

鼠标滚轮缩放图片		  
视频	播放进度	操作
犯罪心理第一季03	[Web网站]已看到35分57秒	续播
2013NCAA联赛 加利福尼亚大学洛杉矶分校VS	[Web网站]已看到69分23秒	续播
[录播]意甲13-14赛季第6轮:AC米兰1-0桑普多利	[Web网站]已看到39分56秒	续播
情妇第一季02	[Web网站]已看到27分55秒	续播
初代吸血鬼第一季03	[Web网站]已看到5分25秒	续播
无耻之徒 第一季 01	[Web网站]已看到2分42秒	续播
疯人疯语第一季04	[Web网站]已看到0分00秒	续播
董浩助阵品牌发布会 感谢陈坤让其再相信爱情	[Web网站]已看到0分00秒	续播

Figure 2-1-3 Watched videos

As shown in the figure2-1-4, the recommendation result is basically the current popular video, which has nothing to do with the viewing history and the interest of the Tencent account bound to the account. It is unable to capture the user's interest, which shows that the recommendation system needs to be improved.



Figure 2-1-4 recommendation result

2.2 International mainstream movie recommendation system

2.2.1 Jinni Movie recommendation system

Jinni can search based on a series of parameters such as the location of the movie's plot time and the award-winning keywords. The search tool used by Jinni is called Movie Genome. Movie Genome is very well-known, and Google TV also uses this tool to achieve personalized search. Jinni's Movie Genome contains 2,200 different parameters, including type background atmosphere, etc. It finds exact parameters to describe a movie based on user reviews and other information. Jinni uses a large number of parameters included in Movie Genome to classify movies and explores user preferences and the degree of association between users.

When searching for a movie in Jinni, Jinni will provide other related movies

based on the related information of the movie. Recommended movies are displayed in image format. The size of the image represents the order of the movies. By default, the order is based on relevance. The larger the movie picture, the stronger the relevance of the search information. Of course, the user can reorder these movies, including the movie's release time and the order of scores. For the recommendation list, the user can adjust the information again to achieve secondary screening. , Information adjustments include movie popularity, movie seriousness, movie reality, etc.

2.2.2 IMDB movie recommendation system

IMDB uses multiple retrievals, that is, multiple keywords are used to filter the demand information at the same time. IMDB can retrieve titles, abstracts, keywords, movie company plots, etc., and also has advanced search capabilities. Advanced search includes advanced title search, advanced name search, and collaborative overlap search. Advanced title search can take into account multiple factors such as the number of votes and ratings of the movie genre, and perform search recommendations.

Advanced name search can search based on factors such as the gender, date of birth, name, and height of the actors in the movie. Collaborative overlapping search can achieve simultaneous retrieval of two titles or two names.

2.2.3 Rotten Tomatoes movie recommendation system

Rotten Tomatoes will track all the comments (some major large movies can reach about 250) and the proportion of positive comments. If the positive comments exceed 60%, the work will be considered fresh; on the contrary, if the positive evaluation of a work is less than 60%, the work will be marked as rotten. There are two types of film evaluation, T-MeterCritics and Top Critics. Scores by professional film critics.

This makes it easy to see the recommended features of RottenTomatoes. The first is a simple answer. You can only choose good or bad. There is no concept of score. The result is based on the percentage of good choices. The advantage of this algorithm is that when the standard is relatively absolute, the likes and dislikes of each vote are obvious and there is no ambiguity. The second is the three-factor scoring method: a movie has 3 freshness, which are rated by ordinary users, professional film critics and famous film critics. The score rated by many professional film critics is the most valuable reference, and it is also Rotten Tomatoes. The most worth watching content.

2.2.4 Movielens movie recommendation system

Movielens provides three ways to search for movies. The first method is title search. The movies obtained by this search will not consider whether users have rated them or whether they have predictions. The second method is genre search, which is searched by the genre of the movie. The genre search will not return

movies that have been rated by the user or movies that have not been predicted. The third method is advanced search, which helps users to improve the search conditions, and it is not a simple title search plus type search. It has different search methods, such as title phrase search date limit search language range search, etc. through advanced search to the first two With the improvement of this basic search method, users can basically realize movie search under fuzzy information.

3 Analysis of Problems in China's Movie Recommendation System

Our project aims to solve the problems in China's movie recommendation market, so the relevant problem analysis is mainly for the Chinese market. With the development of personalized recommendation technology, it cannot avoid inherent problems, such as data sparseness and cold start, and recommendation accuracy. Data sparseness refers to the fact that the total amount of items purchased or rated by general users in the system accounts for about 1% of the total amount of items on the website, resulting in a very sparse evaluation matrix (user item matrix), making it difficult to find the nearest neighboring user set. The similarity calculation cost increases.

From a certain perspective, the cold start problem can be regarded as an extreme case of the sparse problem. It refers to a new item that has not received any user evaluation after the first appearance or a new user has not evaluated any

item after entering the system, resulting in difficult recommendation and poor recommendation accuracy. Therefore, under the premise of ensuring accuracy, reducing the impact of data sparseness and cold-start problems has become a popular direction for personalized recommendations.

For some films, the similarity of user ratings may not be used as the primary factor to recommend. For some experts in film art, he may be more willing to accept recommendations from people with the same professional level as him.. Another more important factor is the credibility of users. Nowadays, there are many untrustworthy things happening on the Internet, such as false ratings and comments. Because many online movie rental companies may hire someone to increase the ratings of a certain movie in order to increase the ratings of a certain movie of their own. This is a problem that the current recommendation algorithm cannot solve.

4 Commercial Value of the Project

4.1 Project technical advantages

4.1.1 Application of tag

Our project applies the characteristics of the tag. In recent years, tags have been widely used in major recommendation websites, making them an effective recommendation tool. Tags are keywords attached to digital objects to describe the objects, but they are not part of the informal classification system. It is a freely

chosen keyword, but it is a simple and powerful tool used to organize search and explore object resources. It has many advantages: First, it more directly reflects the user's feelings about a movie, which is different from the user's Comments, tags do not have a strict organizational structure, but contain rich and clear subject information, thus avoiding the cumbersomeness of user reviews; also, unlike movie type information, tags can store user feelings more flexibly; in addition, tag behavior forms a Based on the three-dimensional relationship of user object tags, in addition to the traditional user-item relationship, the corresponding relationship between item-tag and user-tag is increased. Due to these advantages, tags are becoming more and more popular and widely used in personalized recommendations.

4.1.2 Application of scoring system

The processing of movie scoring in this project refers to the processing method of Movielens. Movielens predicts the user's ratings of other movies based on the user's ratings of some movies. When a new user enters Movielens, it needs to rate at least 15 movies. The rating range is 1 to 5 points, with an interval of 0.5 points. The system's rating of users Perform analysis and feed back the results to users so that users can better understand their own ratings. In the analysis result of Movielens, it will inform users of the top 5 movies rated by other users with the least number of ratings, and the top 5 movies that have not been rated by the user and the highest number of ratings by other users. The user rating is lower

than the top 5 of the average score.

Movies and the top 5 movies with user ratings above average. In addition, the system will also count the number of movies rated by users in each scoring interval and the number of movies on the number of ratings level.

4.2 Commercial Value

The value of the movie recommendation system to video websites is mainly manifested in the following aspects:

- (1) First, the recommendation system improves product intelligence and user experience, and can increase user adhesion.
- (2) A high-quality recommendation system reduces operating costs and improves operating efficiency. With the disappearance of traffic dividends, companies are concerned about reducing costs and increasing efficiency. Traditional portals may require hundreds or thousands of editors for content distribution, which still cannot meet users' fine-grained content needs. The existence of a high-quality recommendation system can realize that several operators are responsible for the distribution of content on a client, and they can do some recommendation interventions on a daily basis.

5 System Design

5.1 Architecture Design

The movie recommendation system adopts a three-layer framework design, which is divided into a presentation layer, a business layer and a data layer. This design pattern achieves the purpose of decentralized attention, loose coupling, logic reuse and standard definition. The overall frame structure of the system is shown in Figure 1.

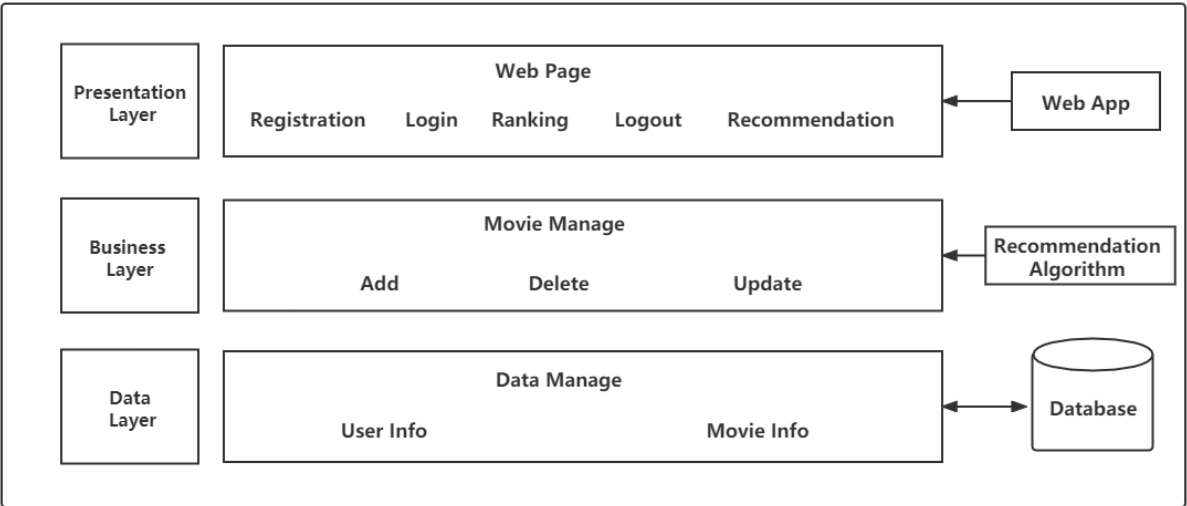


Figure 5-1-5 Watched videos

5.2 Database Design

5.2.1 User Table

Data	Type
UserName	Text
Email	Text

Password	Text
UserId	Integer

5.2.2 Movie Table

Data	Type
MovieName	Text
Poster	Text
IMDBId	Integer

5.2.3 Ranking Table

Data	Type
UserId	Integer
IMDBId	Integer
Ranking	Float

5.3 Recommendation Design

5.3.1 Recommend Process

After the user has scored and submitted the movies he has watched, he can click the recommendation function to get the movie preferences of people who are like his own hobbies. If the movie watched by the user is relatively niche, you can continue to view the recommendations based on the historical record that you have watched. As shown in Figure 1.

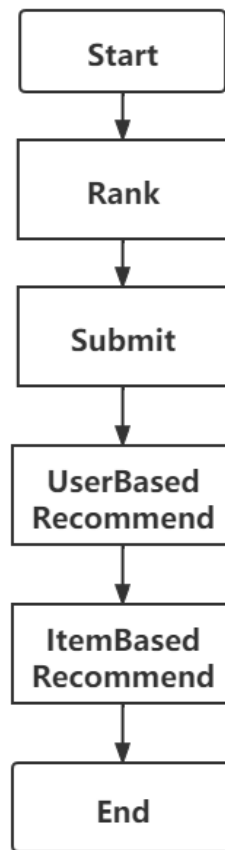


Figure 5-3-6 Watched videos

5.3.2 User-based Collaborative Filtering

User-based collaborative filtering method is shown in the figure. After user A has rated and submitted the movie he has watched, the steps can be as follows:

- (1) The system generates a user-specific matrix for the movies that the user A has watched and rated.
- (2) The system searches for other users with similar interests for user A, and the searching method is to judge the similarity of the two matrices.
- (3) After that, we start to calculate the interest of similar users in movies and take their intersection, so as to find movies that are of common interest to users of this category.

(4) Select K movies from the movies of shared interest as the recommended objects for user A.

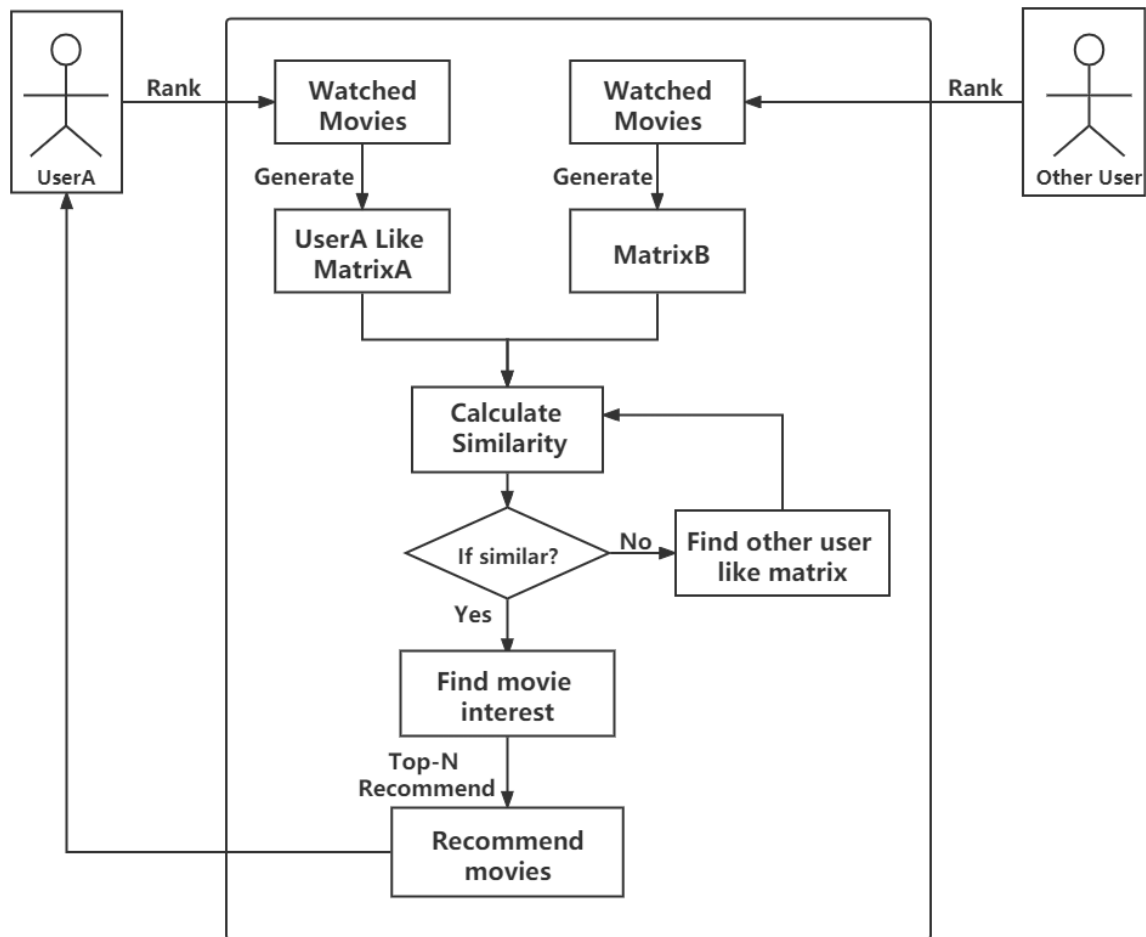


Figure 5-3-2 Watched videos

5.3.3 Item-based Collaborative Filtering

Item-based collaborative filtering method is shown in the figure After user A has rated and submitted the movie he has watched, the steps can be as follows:

(1) If user A gives both movies a high score of 5, it is considered that the two movies have a certain similarity. According to this principle, we can extract movie information from the matrix of user A, and calculate the similarity between movies on this basis.

- (2) Calculate the interest of similar users in movies and take their intersection, so as to find movies that are of common interest to users of this category.
- (3) Select K movies from the movies of shared interest as the recommended objects for user A.

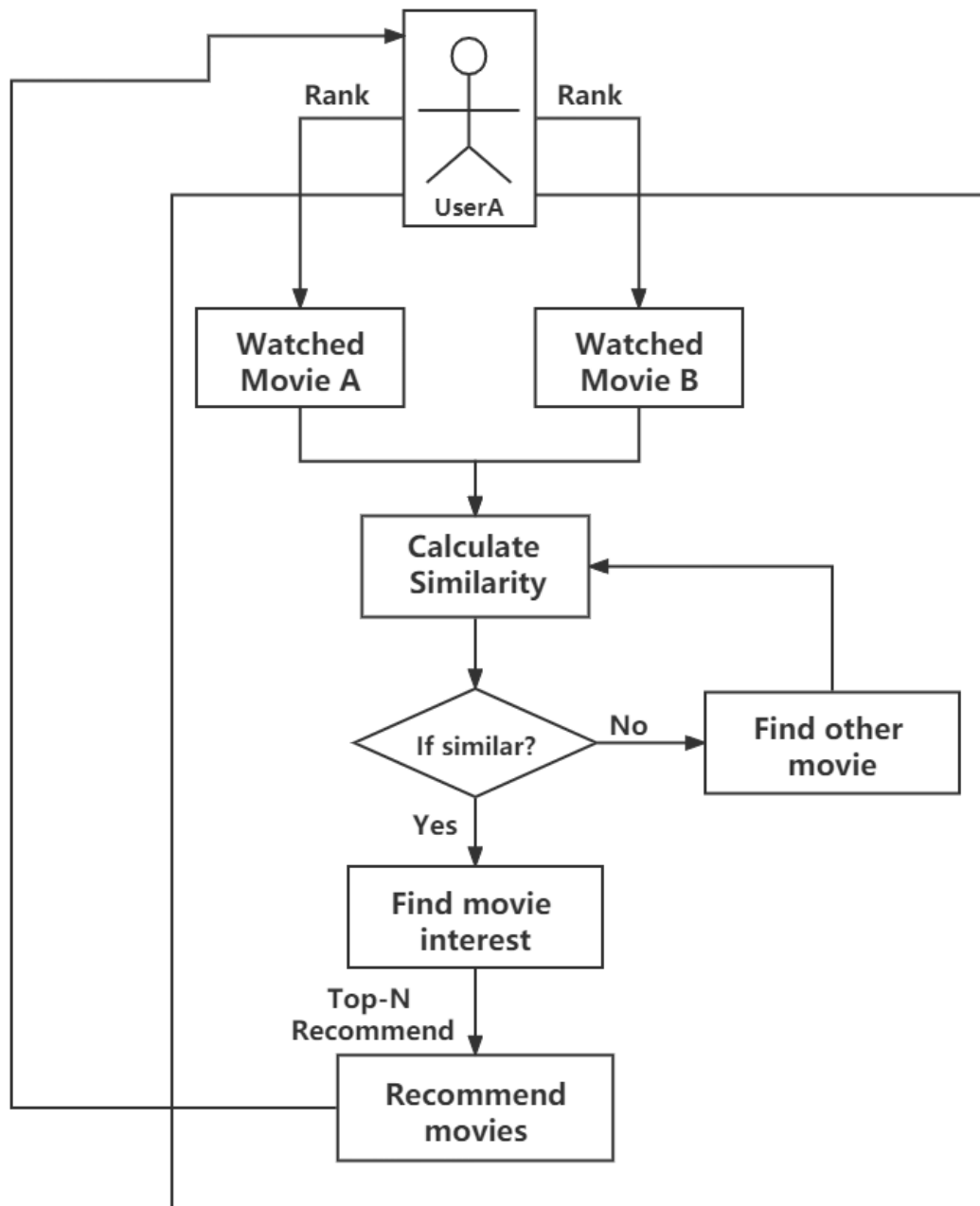


Figure 5-3-3 Watched videos

6 Operating Environment

This project uses the 'python+django+mysql' framework, where the language is mainly python to implement models and algorithms, the database uses mysql5.6, and the visual front-end uses django1.11.

6.1 Introduction of MySQL

MySQL is a small open-source relational database management system. Compared with other large database management systems such as Oracle, DB2, SQL Server, MySQL is small in scale and limited in functions, but it is small, fast, and low in cost. The functions provided are sufficient for slightly more complex applications, and these features make MySQL the most popular open-source database in the world.

MySQL is an open-source relational database management system (RDBMS). The MySQL database system uses the most used database management language-Structured Query Language (SQL) for database management.

Since MySQL is open source, anyone can download it under the General Public License and modify it according to individual needs. MySQL has attracted much attention because of its speed, reliability, and adaptability. Most people think that MySQL is the best choice for managing content when transactional processing is not required.

6.2 Introduction of Django

Django is an advanced Python web framework that can quickly develop secure and maintainable websites. Built by experienced developers, Django takes care of the troublesome part of website development, so you can focus on writing applications without having to re-development. It is free and open source, has an active and prosperous community, rich documentation, and many free and paid solutions.

Django can (and has been) used to build almost any type of website—from content management systems and wikis to social networks and news sites. It can work with any client framework and can provide content in almost any format (including HTML, Rss source, JSON, XML, etc.).

Internally, although it provides options for almost all functions that may be needed (such as several popular databases, template engines, etc.), it can also be extended to use other components if needed.

7 User Guide

7.1 Install project dependencies

In order to ensure the stable operation of the system, the following dependencies are all used stable versions.

(1) Python 3.6+

(2) Django 1.11

(3) MySQL 5.6

7.2 Build Database

(1) Create a database in MySQL and give it a name, such as **`MovieData`**, and create a table **`moviegenre3`** in this database.

```
CREATE TABLE moviegenre3(imdbId INT NOT NULL PRIMARY KEY,title  
varchar(300),poster varchar(600));
```

(2) Import the table with the same name in the folder **'data'** into the database table **`moviegenre3`** and run the command

```
'python manage.py migrate'
```

to build all related tables, and then import the **'users_resulttable'** data in the folder **'data'** into the table with the same name in the database.

7.3 Operating System

(1) Run the command

```
'python manage.py runserver'
```

(2) Click <http://127.0.0.1:8000/> to view the registration, login and scoring pages.