Group Members:

1. Mansoor Shakeel (CS172053)
2. Abdul Samad Vohra (CS162071)

Section: 6A

Movie Recommendation system using Knn algorithm

*Abstract*—In the spread of information, how to quickly find one’s favorite movie in a large number of movies become a very important issue. Personalized recommendation system can play an important role especially when the user has no clear target movie. In this paper, we design and implement a movie recommendation system prototype combined with the actual needs of movie recommendation through researching of KNN algorithm and collaborative filtering algorithm. Then we give detailed principle and architecture of JAVAEE system relational database model. Finally, the test results showed that the system has a good recommendation effect.

# **Introduction**

With the rapid development of Internet technology[1], today's

society has entered the era of Web 2, information overload has

become a reality. How to find the required information in the

mass of data has become a hot research topic. Movie is one of

the main spiritual entertainment, also has the problem of

information overload. In order to solve this problem, this

paper put forward a proposal of personalized movie

recommendation system[1,2].

Personalized recommendation try to know the

characteristics and preferences of the user by collecting and

analysing historical behavior to know what kind of person the

user is, what kind of behavior preference the user has, what

kind of things the user like to share and so on[3.4.5], and

finally understand that user characteristics and preferences

based on the rules of the platform and recommend the

information and goods which the user interested[6.7].

Personalized recommendation system is a kind of

information filtering technology. It is an integrated system

which is a combination of a variety of data mining algorithms

and user related information, to meet the interests or potential

interests of users. The common recommendation system is

categorized as content based recommendation system,

collaborative filtering recommendation system, and hybrid

recommendation system[9,10]. Each recommendation

algorithm has different use range and use condition, it results

in the use of different recommendation algorithm for the

same information recommendation. In the actual application

of recommendation system, the system tends to be a hybrid

recommendation system. That is, to mix the advantage of each

each recommendation algorithm to the recommended process to

effectively improve the recommendation effect. In this paper, the

key research contents is to help users to obtain user-interested

movie automatically in the massive movie information data

using KNN algorithm and collaborative filtering algorithm, and to

develop a prototype of movie recommendation system based on KNN

collaborative filtering algorithm.

## **Literature** **Survey**

•KNN Algorithm

KNN algorithm is called K nearest neighbor classification

Algorithm. The core idea of the KNN algorithm is: if the

Majority of the k most similar neighbors of sample in the

Feature space belongs to a certain category, then the sample is

considered to belong to this category[8]. As shown in Figure

1, the majority of w’s nearest neighbors belong to the x

Category, w belongs to the X category.

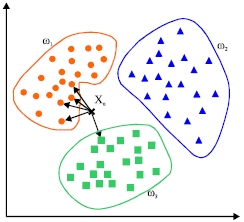


Fig. 1 example of KNN algorithm

•Collaborative filtering algorithm

Collaborative filtering algorithm is categorized as user-based

collaborative filtering algorithm[4] and project-based

collaborative filtering. The basic principles of the two is quite

similar, and this section mainly introduces the user-based

collaborative filtering recommendation algorithm. The basic

idea of collaborative filtering recommendation algorithm is to

introduce the information of similar-interest users to object

users[7]. As shown in figure 2.

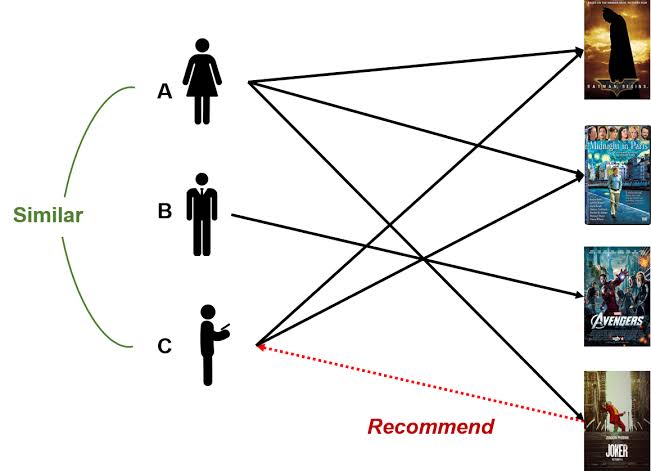


Fig 2 example of UserCF algorithm

User A loves movie A, B, C, and user C likes movie B,D

so we can conclude that the preferences of user A and user C

are very similar. Since user A loves movie D as well, so we

can infer that the user A may also love item D, therefore item

D would be recommended to the user. The basic idea of the

algorithm is based on records of history score of user. Find

the neighbor user as u` who has the similar interest with target

user u, and then recommend the items which the neighbor

user u` loved to target user u, the predict score which target

user u may give on the item is obtained by the score

calculation of neighbor user u` on the item. The algorithm

consists of three basic steps: user similarity calculation,

nearest neighbor selection and prediction score calculation.

### **KNN** **COLLABORATIVE** **FILTER** **ALGORITHM**

KNN collaborative filtering algorithm, which is a

collaborative filtering algorithm combined with KNN

algorithm, use KNN algorithm to select neighbors. The basic

steps of the algorithm are user similarity calculation, KNN

nearest neighbor selection and predict score

calculation[11.12]

## •User similarity computing

The similarity between users is calculated by evaluating the

value of the items evaluated by two users.

Each user uses N dimension vector to represent item score,

for example, to calculate of similarity of U1 and U3, first find

out the set of films that they all scored as {m1, M2, M4, m5}

and relative scores of these films. The score vector of U1 is

{1,3,4,2}, and the score vector of U3 is {2,4,1,5}. The

similarity of U1 and U3 is calculated by the similarity

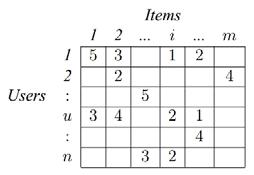
formula[13].

Fig.3 calculation of user's similarity

The similarity between users is calculated by evaluating the

value of the items evaluated by two users.

Each user uses N dimension vector to represent item score,

for example, to calculate of similarity of U1 and U3, first find

out the set of films that they all scored as {m1, M2, M4, m5}

and relative scores of these films. The score vector of U1 is

{1,3,4,2}, and the score vector of U3 is {2,4,1,5}. The

similarity of U1 and U3 is calculated by the similarity

formula[13].

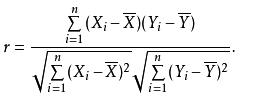
Understand the Relationship Between Pearson Correlation Coefficient and Cosine Similarity

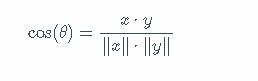
Pearson Correlation Coefficient and Cosine Similarity can

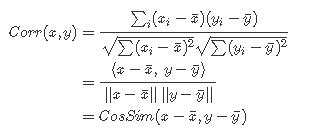
measure the correlation between two varialbes, both of them

are in [-1, 1]. In this tutorial, we will discuss the relationship

between them.

As to pearson correlation coefficient, it is defined as:

As to cosine similarity, it is defined as:

We can find:

It means pearson correlation coefficient is the cosine

similarity between centered versions of \(x\) and \(y\).

•**KNN** **nearest** **neighbor** **selection**

After the calculation of similarity as cosine and pearson between

users,then the algorithm selects a number of users the highest

similarity as the U’s neighbor, denoted as u'. set a fixed value

K for the neighbor selection, select only the most K high

similarity as neighbors regardless of the value of the neighbor

similarity of users. As shown in figure 4.

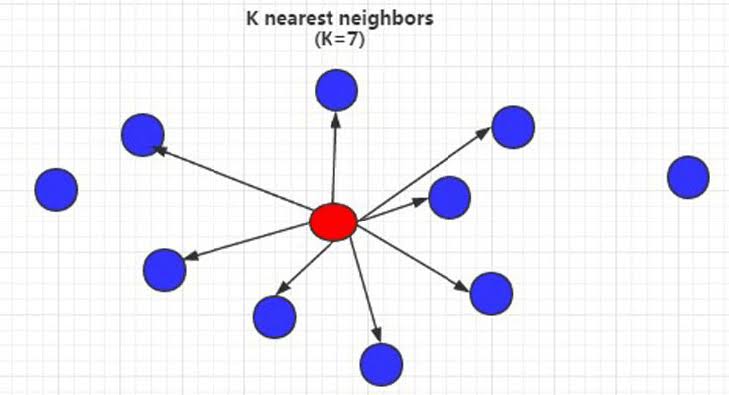
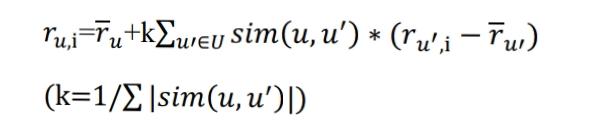


Fig.4 formula of K nearest neighbor when k is 7

Predict score calculation

After determining the user's neighbors, the score can be

predicted according to the score of the neighbor to the item,

The calculation formula is as follows:

*r u, i* used to predict the score of user u to movie i[6].

To sum up, the process of calculating prediction score of

user u for i is as follows:

Step1. Generate user - item two-dimensional matrix of

score as Rmxn, where each score is *ru i..*

Step2: Use principle of cosine similarity or Pearson

correlation similarity to calculate the similarity between each

2 users as *sim (u,u¹)* and generate the user similarity matrix.

Step3: according to the results obtained by Step2, find K

number of score which has the maximum weight, the

corresponding K users is the neighbors of u.

Step4: Use formula 3 to calculate the predictive value of

i for target user u.

In this way, we can calculate the prediction score of the

target users for the non-scored movies, and the N movies with

the highest score can be recommended to the user.

In this paper, KNN collaborative filtering algorithm based

on user is used to implement the recommendation of movie[],

and the collaborative filtering algorithm based on the project

is used to implement the recommendation of the associated

movie[5]. In addition, it can also recommend the movies to

new users according to user registration information, it

can make new and unpopular movie recommendation

according to the film's browsing and score[14].

**Personalized System Design Recommendation**

•Architecture design

The system is based on B/S mode, uses JavaEE architecture,

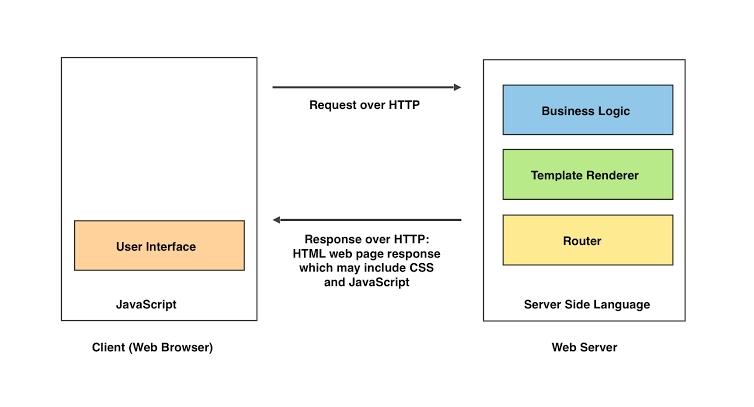
Tomcat server for system deployment, the architecture is shown in figure 5.

Fig.5 System Architecture

Front view is implemented using HTML, CSS,

JAVASCRIPT, the back end uses Struts2, Spring and

Hibernate, the database uses MySQL for storage. The system

is object-oriented to guarantee system of high cohesion and

improve development efficiency using the SSH protocol[17].

Besides, it enhances the maintainability and scalability by

separating Controller layer and View layer to reduce the

degree of coupling between them, making it easier to

maintain and modify the WEB application.

•Database design

Database is the basis of the system, this system uses MYSQL

database, the overall database structure diagram is shown in

the following figure 6, representing the integrity constraints

between the data tables[18].

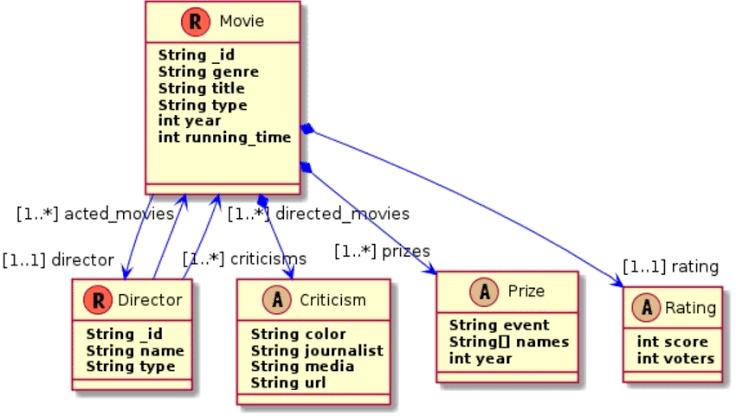


Fig. 6. Relational Model of Database

Table Users is the description of user information, including

user ID, user name, password, registration time, etc. Table

UserSimilar is the description of the user similarity

information, including the user similarity ID, user ID, similar

neighbor user ID, and the value of the similarity. Table Score

is the description of users’ rating information on the film,

which is the direct information source of collaborative

filtering algorithm, it includes the score ID, the user's ID who

give the score ,the value of the score, content of comments.

Table Movie is the description of the movie information,

including the movie ID, movie name, director, movie URL,

etc. Table MovieType is the description of type information

of the movies, including the ID of movies’ type, movie name,

and type ID. Table MovieSimilar is the description of the

movie similarity information, including the movie similarity

ID, movie ID, the ID of highly similar neighbor, the value of

similarity. Both the table UserSimilar and table Moviesimilar

are the basis of the recommendation algorithm and

system[15.16].

### **System Overview**

User registration system will capture the user’s explicit and

implicit behavioral characteristics and these characteristics is

stored in the user database through the user login module.

After logging in the system, the system will make the

appropriate recommendation according to the user’s

information[19.20]. As shown in figure 7.

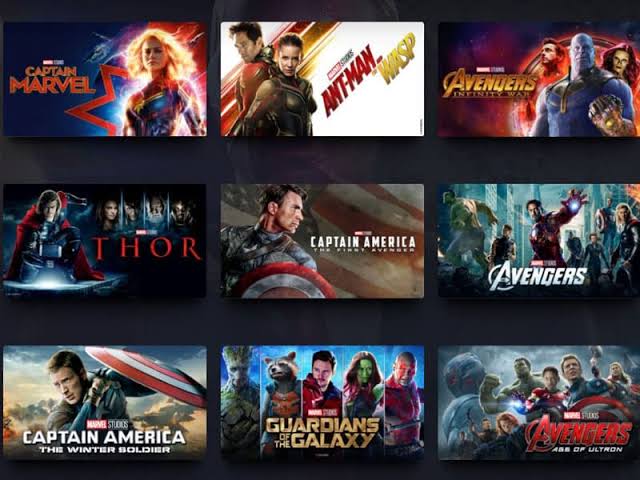


Fig. 7. The display of personal recommendation list

##### CONCLUSION

Under the condition of massive information, the requirements

of movie recommendation system from film amateur are

increasing. This article designs and implements a complete

movie recommendation system prototype based on the KNN

algorithm, collaborative filtering algorithm and

recommendation system technology[10]. We give a detailed

design and development process, and test the stability and

high efficiency of experiment system through professional test.

This paper has reference significance for the development of

personalized recommendation technology.

###### REFERENCE

1.Hu jinmentming. “Application and research of collaborative

filtering in e-commerce recommendation system”, 2010

3rd International Conference on Computer Science and

Information Technology, 07/2010

2. Yan, Bo, and Guanling Chen. “AppJoy : personalized

mobile application discovery”, Proceedings of the 9th

International conference on Mobile systems applications

and services – MobiSys 11 MobiSys 11, 2011.

3.K. Liao. Prototyping a recommender system step by step

Part 1: KNN item-based collaborative filtering. [Online]. Avail

Able: <https://towardsdatascience.com/prototyping-a-recommender-system->

Step-by-step-part-1-knn-item-based-collaborative-filtering-637969614ea

4. R. Jain, “Rohit9314/movie-recommendation-system,”

Original-date: 2018-06-23T09:10:20Z. [Online]. Available:

<https://github.com/Rohit9314/Movie-Recommendation-System>

5. Okkalioglu, M., Koc, M., Polat, H.: On the discovery of

fake binary ratings. In: Proceedings of the 30th Annual

ACM Symposium on Applied Computing, SAC 2015,

pp. 901–907. ACM, USA

6. Kaleli, C., Polat, H.: Privacy–preserving naïve Bayesian

Classifier based recommendations on distributed data.

Comput.Intell.31(1).47–68(2015). <http://dx.org/10.1111/coin.12012MathSciNet>

CrossRefGoogle Scholar

7. Polat, H., Du, W.: Privacy-preserving top-n

Recommendation on distributed data. J. Am. Soc. Inf. Sci.

Technol. 59(7), 1093–1108

(2008). <http://dx.org/10.1002/asi.20831CrossRefGoogle>

Scholar