Movie Recommendation System

Elif Özcan, Büşra Küden
Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi, 34220 Istanbul, Türkiye
{elfozcn08, busrakuden}@gmail.com

Özetçe* —Bu belge, SİU 2017 bildirisi hazırlamanız için bir taslak içermektedir. Bu sebeple lütfen taslaktaki başlık, özet ve diğer format stillerini kullanınız. **Dikkat: Bildiri Başlığında ve Özetlerde Sembol, Özel ve Matematiksel Karakterler kullanmayınız.

Anahtar Kelimeler—doküman biçimi, stil, anahtar kelimeler.

Abstract*—This electronic document is a "live" template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. **CRITICAL: Do Not Use Symbols, Special Characters, or Math in Paper Title or Abstract.

Keywords—component, formatting, style, styling, insert (key words).

I. METHOD TEST AND RESULTS

There are 610 users in our data set and there are movies that they all rate. We deleted 10 rates for each user in the data set for measure the success of IBCF(Item Based Collaborative Filtering) and UBCF(User Based Collaborative Filtering) methods. For each film in the created test data set, we calculated estimated rate using methods. The error between the estimated rate and the original rate in the test data set was calculated. When calculating the error, we used mean squared error method. We calculated the average error by dividing the calculated total error by the number of predictions made. We used a different data set for the CBR(Content Based Recommendation) method. Since the data set we used did not have any user rates, we did not perform the test by estimating the votes.The applied method is described in detail under the heading Content Based Recommendation.

$$accuracy = 100 - error * 20 \quad (1)$$

Since the calculated average error is out of 5, we used the equation (1) to find the percentage of success.

A. Item Based Collaborative Filtering

For testing IBCF method ,for each film in the created test data set, we calculated estimated rate using IBFC method. Then we calculated the error with the results we obtained.

1) Item Based Collaborative Filtering Test Results: We reduced number of users in data set and tested our data set to see the results of the method in different data sets. We tested the IBCF method using data sets containing 100, 305 and 610 users. The number of movies and test results in the data sets are given in Table 1.

Table 1 TEST RESULTS OF ITEM BASED COLLABORATIVE FILTERING METHOD WITH 3 DATA SETS

Number of movies	Error out of 5	Accuracy
15448 movies	0,806	%83,87
46621 movies	0,787	%84,25
100836 movies	0,809	%83,80

B. User Based Collaborative Filtering

For testing UBCF method ,for each film in the created test data set, we calculated estimated rate using UBFC method. Then we calculated the error with the results we obtained.

1) User Based Collaborative Filtering Test Results: As we did when testing the IBFC method, we reduced number of users in data set and tested our data set to see the results of the method in different data sets. We tested the UBCF method using data sets containing 100, 305 and 610 users. The number of movies and test results in the data sets are given in Table 2.

Table 2 TEST RESULTS OF USER BASED COLLABORATIVE FILTERING METHOD WITH 3 DATA SETS

Number of movies	Error out of 5	Accuracy
15448 movies	1,543	%69,13
46621 movies	1,375	%72,49
100836 movies	1,499	%70,00

C. Content Based Recommendation

Since the data set we used for the content based recommendation system is not suitable for testing, we have adapted the data sets we use for the collaborative filtering method and tested the system success. First, we found the total number of votes and the average of the votes of each user, so that the votes, genres, user ids, movie titles and ids of the movie that each user receives in the first 10 votes are divided into test data set.Then, CBR method is tested with the created data set.

1) Content Based Recommendation Test Results: We divided the numbers of recommended movies whose rates are above the average rates by number of movies rated for calculating accuracy. As we did when testing the IBFC and UBCF methods, we reduced number of users in data set and tested our data set to see the results of the method in

different data sets. We tested the CBR method using data sets containing 100, 305 and 610 users. The number of movies and test results in the data sets are given in Table 3.

Table 3 TEST RESULTS OF CONTENT BASED RECOMMENDATION METHOD WITH 3 DATA SETS

Number of movies	Accuracy
15448 movies	%59,30
46621 movies	%60,25
100836 movies	%58,49

D. Conclusion of Tests

Before performing the tests, we knew that even if both methods seem equal, the item based method will be more successful than the user based method. The reason why the item based method is more successful is that items can be evaluated more easily than the users. While it can express movies with a few specific features, users may have different and strange tastes. For this reason, the item-based approach is more meaningful than the user-based approach, and with our tests, we have shown that this information is also correct for our data set.

Even if the success rate of the CBR method is lower than the Collaborative Filtering methods, we added a module that works with the content based method to the system as the content based recommendation method can use it as a solution to the cold start problem.

In our tests with different numbers of users, we could not achieve very meaningful results in three methods.

REFERENCES