Volume 4, Issue 7, July 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Hybrid Recommender System under Temporal Vector

Sarita Dhankhar
Pursuing M.Tech
CBS Group of Institution
Jhajjar, Haryana

Neha
Assistant Professor
CBS Group of Institution
Jhajjar, Haryana

Abstract: This paper describes a hybrid temporal effective recommendation system. Recommender system is an intelligent data mining approach that actually performs the prediction for new data values based on existing dataset analysis. Recommender systems have become extremely common in recent years and are applied in a variety of applications. The most popular ones are probably movies, music, news, books, research articles, social tags and products in general. In this work, a recommender system is suggested for movie interest analysis for specific user group. In this work, the recommendation will be applied on user side as well as movie side. Hybrid recommendation system will combine the functionality of content based as well as collaborative recommender system. At the later stage, a weighted approach will be defined on content based and collaborative methods to perform the recommendation. This two level hybrid approach will give more accurate results.

Keywords- Recommender system; content based filtering; collaborative filtering; hybrid approach; group similarity

I. INTRODUCTION

People use recommendations from others in everyday life. Conversations, News, magazines belonging to different products or services are just some examples of the ways people can get to know about others' opinions. Recommender systems use this social fact to help people find valuable information about plenty of books, restaurants, movies, web pages, foods, and any kind of products or services available on internet. Recommendation systems are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that user would give to an item. Recommender system use different methods for preparing recommendations.

In this work, we propose a hybrid system for movie recommendations that mediates the data sparsely problem and reduce the data rate. The presented work is based on hybrid recommendation system so that more accurate results are expected from the system. Other significance of this system is that it is based on temporal vectors that will improve the accuracy in prediction of movie rank.

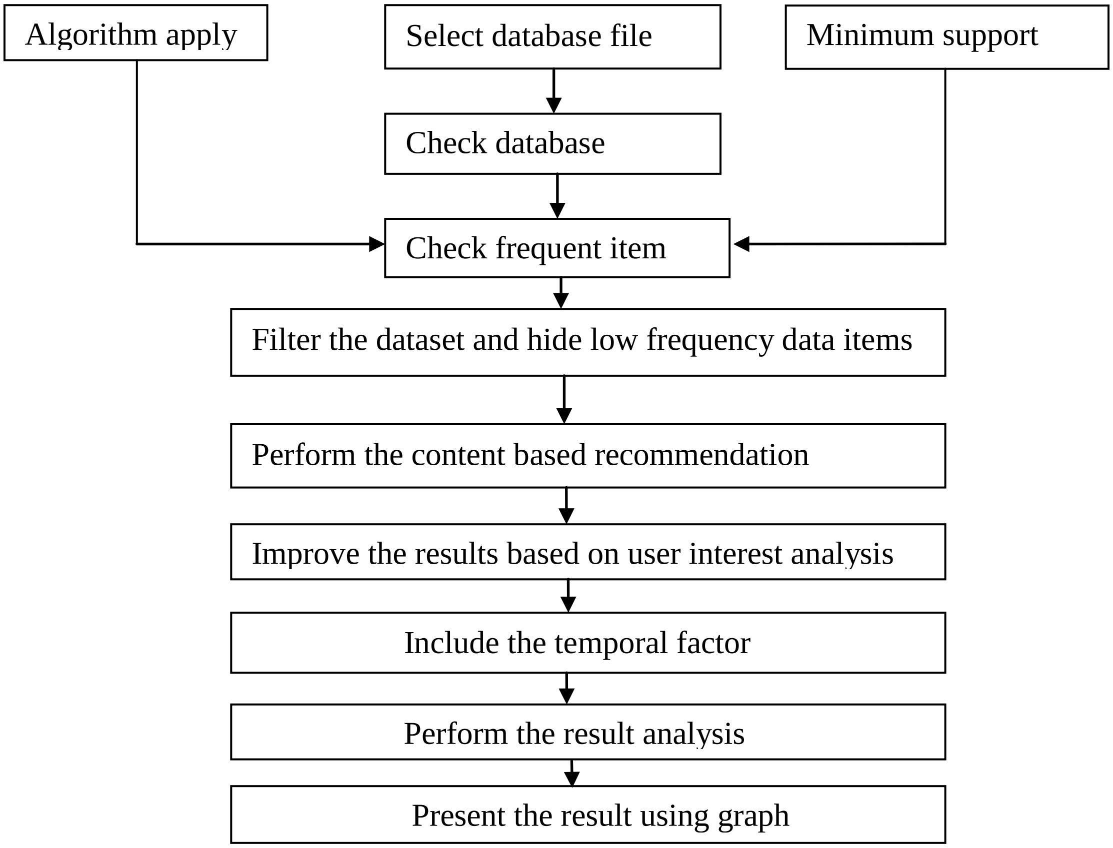
II. PROBLEM DEFINITION

The recommender system is the analytical study on the identification of user interest without getting any previous input from the user. The interest analysis is here performed by getting the user information and performing the intelligent analysis. In this work, the user similarity analysis is performed to identify the similar user on which the recommendation can be applied. In this present work a hybrid model is defined using the content based and collaborative mining for movie rank prediction. The dataset defined for the process will have three main tables called user data, movie data and rank data. The presented work will first perform content based analysis on movie data and user data separately to identify the user category and movie category. Later the collaborative analysis on dataset will be performed to identify the expected movie rank for the input user.

III. OBJECTIVE

- The main objective of work is to design a weighted hybrid recommender system model using two level content based and one level collaborative analysis.
- The objective of the work is to perform the similarity analysis between users and movies using weighted approach.
- The objective of the work is to define an effective statistical formula for effective matching of records.
- Include the temporal factor in the model to derive more accurate results from the system.

IV. ROPOSED METHODOLOGY



V. SIMILARITY ANALYSIS

The recommender system is about to perform the similarity based analysis. This analysis is required to performed at all three levels for all the available three tables. The first level is performed on user database to perform the user similarity analysis. In second stage, the analysis will be performed on movie database for movie similarity analysis. In third stage, the ranking similarity analysis is performed. To perform these three level analysis a hybrid model is presented. In which two stage model is presented. At first stage, the content based analysis is performed and in second stage, the collaborative analysis is performed for similarity analysis.

Attribute Similarity

The effectiveness of the recommender system is based on the User attribute similarity. The attribute of this dataset includes the gender, age, occupation and location. The similarity analysis between two users is shown here

$$sim_D(u_1, u_2) = \frac{\sum_{f \in F} w_f \, X \, sim(u_{1f} - u_{2f})}{\sum_{f \in F} w_f}$$

where *f* represents a feature of the user from the set of all demographic features *F*, *w* represents the relative weight of feature *f*,

u_{1f} and u_{2f} represent the values of *f* for u_1 and u_2 and

$Sim(u_{1f}, u_{2f})$ represents the similarity between values of *f* for u_1 *and* u_2

Rating Similarity

The rating formula is given here under

$$u_{ij} = \frac{\sum_{x \in P_i \cap P_j} (r_{ix} - \overline{r_i})(r_{jx} - \overline{r_j})}{\sqrt{\sum_{x \in P_i \cap P_j} (r_{ix} - \overline{r_i})^2 \sum_{x \in P_i \cap P_j} (r_{jx} - \overline{r_j})^2}}$$

Where $\overline{r_i}$ is the average rating user i give to all items and $\overline{r}$ is the average rating user j give to all items, r_{ix} is the rating given by user i to item x and r_{jx} is the rating of j to item x.

Item Rating similarity between item x and y is given by IB-PCC as

$$p_{xy} = \frac{\sum_{i \in U_x \cap U_y} (r_{ix} - \overline{r_x})(r_{jx} - \overline{r_y})}{\sqrt{\sum_{i \in U_x \cap U_y} (r_{ix} - \overline{r_x})^2 \sum_{i \in U_x \cap U_y} (r_{iy} - \overline{r_y})^2}}$$

Where $\overline{r}$ is the average rating given to item x and $\overline{r}$ is the average rating of item y.

5.5 Rating Prediction

Proposed approach is a hybrid approach that includes the indirect similarities between items as well as user.

Attribute similarities include the feature similarities between users and items which serve as contributing information in prediction. For the calculation, demographic features of users are to be weighted according to their importance. So with reference to [28], the weights assigned to the features are as:

$W_{age} = 0.41$
$W_{gender} = 0.56$
Woccupation = 0.03

To calculate demographic similarity between users, a function is designed whose prototype is:

Function similarityuser1= similarityusers (UserIdi, UserIdj, NAge, Occupation, Gender)

Item Attribute Similarity is calculated in as similar way with the following weight values as suggested in [24]:

$W_{genre} = 0.04$
$W_{date} = 0.38$
Wcountry = 0.07
Wcompany = 0.22

Formula used is:

similaritySite1=ratio.04+datesim*.38+Companysim*0.22+Countriessim*0.07*

where, ratio gives the similarity between the Sites on basis of genre and other terms are self explanatory like companysim is for similarity on basis of production company and so on.

5.7 MAE Calculation

The steps used in evaluation of results are:

1. The dataset is divided in terms of training and testing datset, the size of training dataset is considered 90% of the original set and testing dataset size is 10% of original dataset
2. This method is called the 10-fold dataset and the work is tested on different testing sets under 10 fold method.
3. The error estimation is here defined in terms of mean square error. The mean square error is defined the ratio difference between the actual rating value and the predicted rating for particular record. The formula for mean square error is shown here under

$$M_s = \frac{1}{|D_s|} \sum_{r_{ix} \in D_s} |r_{ix} - \hat{r}_{ix}|$$

Here

rix is defined as the actual dataset rating
$\hat{r}_{ix}$ is defined as the predicted dataset rating

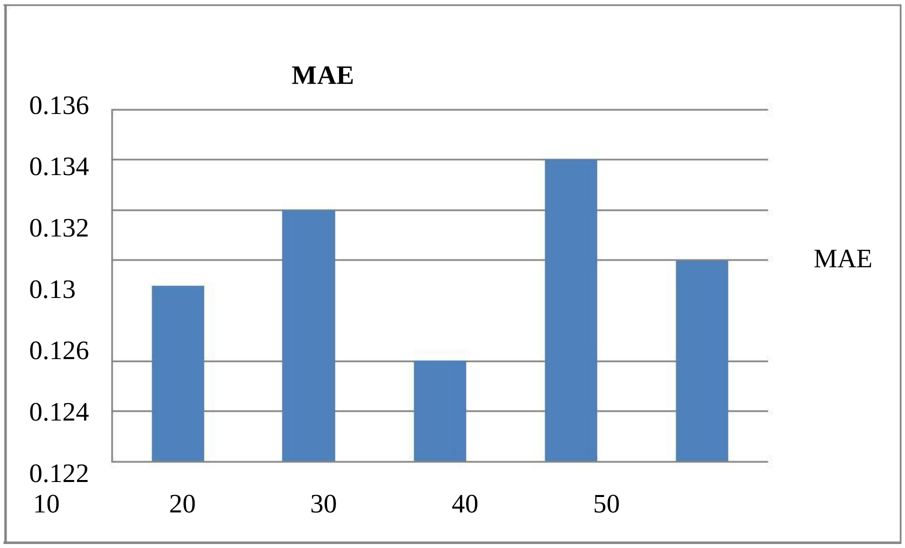
4. To obtain the overall error over the dataset, the sum of all the ratio error difference is considered.

Here the work is defined for different datasets and the error rate is recorded for each set. The error estimation is defined here as under

Table 5.1: MAE based Error Analysis

Method	Mean Absolute Error
10 elements dataset	.129
20 elements dataset	0.132
30 elements dataset	0.126
40 elements dataset	0.134
50 elements dataset	0.130

The results obtained from the work are shown in the form of figure 5.3



MAE Dataset

Here figure is showing the error rate estimation over the dataset. Here x axis shows the number of test records and y axis shows the MAE.

VI. CONCLUSION

In this present work we have defined a recommender system to identify the ranking assigned to a movie. In this work, we have represented a hybrid model to perform the analysis. The hybrid model performed the analysis based on content based as well as the collaborative filtering. To perform this we have taken an authenticated dataset with three tables. One for user ,second for site and third for ranking. Now while predicting the ranking of a new user. At first the content based similarity match is performed to identify the similar users in the dataset. In the second stage, the rank provides by same kind of users is analyzed under the collaborative filtering and at the final stage the collective decision is taken regarding the site rank. The dimension included in this work is the temporal factor. It means instead of analyzing the rank of all similar users a time based range is setup in this work. The presented work is analyzed under the error rate. The obtained results show that the presented work is effective enough to provide correct results.

VII. FUTURE SCOPE

The present work can be improved under different dimensions.

1. The main work can be done in future to reduce the error rate. In this work we have used 3-4 parameters while performing the content based match. More attributes can be considered to obtain more accurate results.
2. The another work can be done in same area with different datasets

REFERENCES

[1] Kleanthi Lakiotaki," UTA-Rec: A Recommender System based on Multiple Criteria Analysis", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp 219-225)

[2] Juan A. Recio-García," Prototyping Recommender Systems in jCOLIBRI", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp243-250)

[3] Paul Resnick," The Influence Limiter: Provably ManipulationResistant Recommender Systems", RecSys'07, October 19–20, 2007, Minneapolis, Minnesota, USA. ACM 9781595937308/07/0010 (pp 25-32)

[4] Linas Baltrunas," Exploiting Contextual Information in Recommender Systems", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp 295-298)

[5] Erich Christian Teppan," Implications of Psychological Phenomenons for Recommender Systems", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp 323-326)

